A COURSE OF HIGHER MATHEMATICS

V. I. Smirnov

Volume III Part 1

LINEAR ALGEBRA

INTERNATIONAL SERIES OF MONOGRAPHS IN

PURE AND APPLIED MATHEMATICS

GENERAL EDITORS: I. N. SNEDDON, M. STARK AND S. ULAM

VOLUME 59

A COURSE OF
HIGHER MATHEMATICS

III/1

LINEAR ALGEBRA

A COURSE OF
Higher Mathematics

VOLUME III
PART ONE

V. I. SMIRNOV

Translated by

D. E. BROWN

Translation edited by

I. N. SNEDDON
Simson Professor in Mathematics
University of Glasgow

PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
PARIS · FRANKFURT
1964

PERGAMON PRESS LTD.
Headington Hill Hall, Oxford
4 & 5 Fitzroy Square, London W. 1

PERGAMON PRESS (SCOTLAND) LTD.
2 & 3 Teviot Place, Edinburgh 1

PERGAMON PRESS INC.
122 East 55th Street, New York 22, N. Y.

GAUTHIER-VILLARS ED.
55. Quai des Grands-Augustins, Paris 6

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt am Main

U. S. A. edition distributed by
ADDISON-WESLEY PUBLISHING COMPANY INC.
Reading, Massachusetts · Palo Alto · London

Library of Congress Catalog Card Number 63-10134

This translation has been made from the Russian Edition of
V. I. Smirnov's book *Курс высшей математики (Kurs vysshei matematiki)*,
published in 1957 by Fizmatgiz, Moscow

MADE IN GREAT BRITAIN

CONTENTS

INTRODUCTION

A BRIEF account of the history of this five-volume course of higher mathematics has been given in the Introduction to Vol. I of the present English edition. This volume and the subsequent ones were, from the first Russian edition (1933), entirely the responsibility of Professor Smirnov.

In most texts on the methods of mathematical physics algebraic methods play a minor role compared with methods based on the theory of functions. This is not so in Professor Smirnov's scheme. In this first part of Vol. III a full account is given of the two branches of modern algebra — linear algebra and the theory of groups — which are most frequently used in theoretical physics. There is a detailed treatment of the theory of determinants and matrices and of quadratic forms including all the results necessary for an understanding of the concepts of functional and Hilbert space. The second part is devoted to a full account of the basic theory of groups and of the linear representations of groups. Novel, in a first course on algebra, is the inclusion of the elements of the theory of continuous groups.

This volume is quite obviously of interest to applied mathematicians and theoretical physicists but its claims as providing material for a first course in abstract algebra for students of pure mathematics should not be disregarded.

I. N. SNEDDON

PREFACE TO THE FOURTH RUSSIAN EDITION

In the present edition the third volume has been divided into two parts in connection with the addition of new material. The first part contains all material referring to linear algebra, to the theory of quadratic forms, and to the theory of groups. I was greatly assisted in compiling the additional material by D. K. Faddeyev. He was partly responsible for the clarification of the simplicity of rotation and Lorentz groups, for the presentation of the material referring to the formation of groups with given structural constants and to integration over groups [70, 81, 87, 88, 89, 90]. I am very grateful to him for his assistance.

V. SMIRNOV

CHAPTER I

DETERMINANTS. THE SOLUTION OF SYSTEMS OF EQUATIONS

§ 1. Properties of determinants

1. Determinants. We shall start the present section by taking the simple algebraical problem of the solution of a system of first degree equations. This will lead us to the important concept of determinant.

We first consider some simple, particular cases. A system of two equations with two unknowns may be written:

$$a_{11}x_1 + a_{12}x_2 = b_1,$$
$$a_{21}x_1 + a_{22}x_2 = b_2.$$

The coefficients a_{ik} of the unknowns are distinguished by two subscripts, the first indicating the equation in which the coefficient occurs, and the second showing with which unknown it is associated.

We know that the solution of the system is

$$x_1 = \frac{b_1a_{22} - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}}; \quad x_2 = \frac{a_{11}b_2 - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}.$$

We next take three equations with three unknowns:

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1,$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2,$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3,$$

the same notation as above being used for the coefficients. We rearrange the first two equations as

$$a_{11}x_1 + a_{12}x_2 = b_1 - a_{13}x_3,$$
$$a_{21}x_1 + a_{22}x_2 = b_2 - a_{23}x_3.$$

We solve these with respect to the unknowns x_1 and x_2 in accordance with the previous formula:

$$x_1 = \frac{(b_1 - a_{13}x_3)\, a_{22} - a_{12}\,(b_2 - a_{23}x_3)}{a_{11}a_{22} - a_{12}a_{21}};$$

$$x_2 = \frac{a_{11}(b_2 - a_{23}x_3) - (b_1 - a_{13}x_3)\, a_{21}}{a_{11}a_{22} - a_{12}a_{21}}.$$

On substituting these expressions in the last equation, we obtain an equation for the unknown x_3; solution of this latter equation leads us to the final expression for this unknown as

$$x_3 = \frac{a_{11}a_{22}b_3 + a_{12}b_2a_{31} + b_1a_{21}a_{32} - a_{11}b_2a_{32} - a_{12}a_{21}b_3 - b_1a_{22}a_{31}}{a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}}. \quad (1)$$

We must carefully examine the construction of this expression. We note first of all that the numerator can be obtained from the denominator simply by replacing the coefficients a_{i3} of the unknown to be determined by the constant terms b_i. This now leaves us with the elucidation of the rule for forming the denominator, which contains no constant terms and is made up solely of coefficients of the system. We write down the coefficients in the form of a square array, which preserves the order in which they appear in the system:

$$\begin{Vmatrix} a_{11}, & a_{12}, & a_{12} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{Vmatrix}. \quad (2)$$

Our array consists here of three rows and three columns; the numbers a_{ik} are known as its elements. The first subscript shows the row in which the element appears, and the second subscript the column. We now write out the denominator of (1):

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \quad (3)$$

It can be seen to consist of six terms, each of which is the product of three elements of array (2), with one element taken from each row and one from each column. The products have the form:

$$a_{1p}a_{2q}a_{3r}, \quad (4)$$

where p, q, r are the integers 1, 2, 3 arranged in some definite order. Thus, the second, as well as the first subscripts form a set of the

integers 1, 2, 3, and to obtain all the terms of expression (3), we have to take all the possible orders of the second subscripts p, q, r in (4). There are clearly six possible permutations of the second subscripts:

$$1, 2, 3; \quad 2, 3, 1; \quad 3, 1, 2; \quad 1, 3, 2; \quad 2, 1, 3; \quad 3, 2, 1, \tag{5}$$

with the result that we obtain all six terms of the expression (3). But some products (4) appear with the plus sign in expression (3) and others with the minus sign, so that we finally have to indicate some rule for the choice of sign. We notice that the products (4) with the plus sign have second subscripts forming the following permutations:

$$1, 2, 3; \quad 2, 3, 1; \quad 3, 1, 2, \tag{5_1}$$

whilst the products with the minus sign have second subscripts in the permutations:

$$1, 3, 2; \quad 2, 1, 3; \quad 3, 2, 1. \tag{5_2}$$

We now indicate how permutations (5_1) differ from permutations (5_2). We refer to the fact that a larger number comes in front of a smaller as an inversion in the permutation, and we calculate the number of inversions in permutations (5_1). There is no inversion in the first permutation, i.e. the number of inversions is zero. We pass to the second permutation and compare the magnitude of each number appearing in it with all those that follow. We see that there are two inversions here: the 2 comes in front of the 1, and the 3 comes in front of the 1. It may readily be seen in the same way that the third of permutations (5_1) contains two inversions. In short, all the permutations (5_1) can be said to contain an even number of inversions. On carrying out a similar investigation of permutations (5_2), we see that they all contain an odd number of inversions. We are now able to formulate a rule of signs in expression (3): the products (4) appear in (3) without change where the number of inversions in the permutation formed by the second subscripts is even. In contrast, the products appear in expression (3) with the minus sign when the permutation formed by the second subscripts contains an odd number of inversions. Expression (3) is known as a determinant of the third order, corresponding to the array of numbers (2). The above discussion can now be easily generalized for the case of a determinant of any order.

Suppose we have n^2 numbers, arranged in a square array with n rows and n columns:

$$\left\| \begin{matrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} \end{matrix} \right\| \tag{6}$$

The elements a_{ik} of this array are given complex numbers, whilst the subscripts i and k indicate that the number a_{ik} stands at the intersection of the ith row and kth column. We form all the possible products of elements of array (6) such that they contain one element from each row and one from each column. These products will have the form

$$a_{1p_1} a_{2p_2} \ldots a_{np_n}, \tag{7}$$

where $p_1, p_2, \ldots, p_n$ are the numbers $1, 2, \ldots, n$, arranged in a certain order. To obtain all the possible products of form (7), we have to take all the possible permutations of the second subscripts. We know from elementary algebra that the number of these permutations is equal to factorial n:

$$1 \cdot 2 \cdot 3 \ \ldots \ n = n!$$

Each permutation will have a certain number of inversions compared with the original permutation

$$1, \ 2, \ 3, \ \ldots, \ n.$$

The products (7) with second subscripts, forming a permutation with an even number of inversions, are taken without change, whereas we write a minus sign in front of the products in which the permutation of second subscripts has an odd number of inversions. The sum of all the products thus obtained is called an nth order determinant, corresponding to the array (6). This sum will evidently contain $n!$ terms. The definition we have given may readily be expressed as a formula. We shall use the following notation. Let p_1, $p_2, \ldots, p_n$ be a permutation of the numbers $1, 2, \ldots, n$. We denote the number of inversions in this permutation by the symbol

$$[p_1, p_2, \ \ldots, \ p_n].$$

Then the definition given above of the determinant, corresponding to array (6), can be expressed as follows, the array being written

between vertical lines in order to indicate the determinant:

$$\begin{vmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & \dots, & a_{nn} \end{vmatrix} = \sum_{(p_1, p_2, \dots, p_n)} (-1)^{[p_1, p_2, \dots, p_n]} a_{1p_1} a_{2p_2} \dots a_{np_n}. \tag{8}$$

The summation extends over all the possible permutations of the second subscripts, i.e. over all the possible permutations $(p_1, p_2, \dots, p_n)$. When referring to the array as such, and not to the determinant formed from it, we write it between double vertical lines.

It may be noticed that the factors in each product in expression (3) have been arranged so that the first subscripts form the basic permutation 1, 2, 3, and hence all our remarks have been concerned with the permutations formed by the second subscripts. Instead, we can write the factors in the products so that the second subscripts always appear in increasing order, and (3) becomes in this case:

$$a_{11}a_{22}a_{33} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13}. \tag{9}$$

The first subscripts here give all the possible permutations p, q, r, and exactly the same rule of signs as above can be stated for the terms of (9), though now with respect to the first subscripts. This leads us to consider, along with sum (8), the analogous sum:

$$\sum_{(p_1, p_2, \dots, p_n)} (-1)^{[p_1, p_2, \dots, p_n]} a_{p_1 1} a_{p_2 2} \dots a_{p_n n}. \tag{10}$$

This latter sum clearly consists of the same terms as sum (8). We shall see later that its terms have the same signs as in (8), i.e. sum (10) coincides with sum (8), as in the case $n = 3$.

We go back finally to the case $n = 2$. Here the array has the form

$$\left\| \begin{matrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \end{matrix} \right\|$$

and (8) gives the following expression for the second order determinant corresponding to the array:

$$\begin{vmatrix} a_{11}, & a_{12} \\ a_{21}, & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

It is clear from the above that an account of the properties of determinants requires a closer acquaintance with the properties of permutations and these form the subject of our next section.

2. Permutations. Suppose we have any n elements arranged in a definite order. This is referred to as a permutation of the elements. We prove first of all that there are $n!$ different permutations. This is obvious with $n = 2$, since two elements can give two different permutations. With $n = 3$, the result follows directly from the list of permutations (5), where the numbers 1, 2, 3 play the part of elements; we can easily verify that (5) gives all the possible permutations of these three elements. We prove our assertion for any integer n by induction. We assume that the assertion is true for a given n and show that it is then valid for $(n + 1)$ elements. Thus, having assumed that n elements give $n!$ permutations, we consider $(n + 1)$ elements which we shall write as

$$C_1, C_2, \ldots, C_{n+1}.$$

We start by considering the permutations in which C_1 is the first element. In order to obtain all these permutations, we must write C_1 in the first position, then write down all the possible permutations of the remaining n elements. The number of these latter permutations is equal to $n!$ by hypothesis, and hence, the number of permutations of elements C_k starting with C_1 is equal to $n!$ Similarly, the number of permutations of elements C_k starting with C_2 is likewise equal to $n!$ In general, the number of different permutations of elements C_k will be altogether

$$n! \cdot (n + 1) = 1 \cdot 2 \cdot \ldots n. \ (n + 1) = (n + 1)!,$$

which is what we required to show.

We can naturally assume that our elements are taken as the integers starting from unity, and we shall confine ourselves to this case in future. We define a *transposition as an operation in which the positions of two elements in a permutation are interchanged.* It follows at once that we can obtain from a given permutation any other permutation by carrying out a certain number of transpositions. For instance, let us take the two permutations of four elements

$$1, 3, 4, 2; \quad 2, 4, 1, 3.$$

We can pass from the first of these permutations to the second with the aid of the following series of transpositions:

$$1, 3, 4, 2 \rightarrow 2, 3, 4, 1 \rightarrow 2, 4, 3, 1 \rightarrow 2, 4, 1, 3.$$

Three transpositions have been needed here in order to pass from the first permutation to the second. If we had used different trans-

positions, our passage from the first permutation to the second would have been by means of a different series; in other words, the number of transpositions required for passage from one permutation to another is not strictly defined. The essential fact that we want to show is that the different numbers of transpositions that may be used are either all even or all odd for two given permutations. This may be explained by bringing in the idea of an inversion which we used in the previous article. Let us take permutations of the n elements $1, 2, \ldots, n$. We call

$$1,\ 2,\ \ldots,\ n, \tag{12}$$

the *normal order*, where the numbers appear in increasing order. We say that there is an inversion in a given permutation when two elements appear in it in a different order to that which they have in the normal order, in other words, when a larger number comes on the left of a smaller number. We define *even permutations as those in which there is an even number of inversions,* whilst *odd permutations are those where the number of inversions is odd.* The following theorem is fundamental for what follows.

A transposition changes the number of inversions by an odd number.

We take the permutation

$$a,\ b,\ \ldots,\ k,\ \ldots,\ p,\ \ldots,\ s \tag{13}$$

and suppose that the elements k and p are transposed. After the transposition, the arrangement of k and p with respect to the elements to the left of k and the right of p remains as before. The only change is in regard to the elements of the permutation lying between k and p, except, of course, that the arrangement of k and p with regard to each other is likewise changed. Let us work out the total change in the number of inversions. Let there be altogether m elements lying between k and p in permutation (13), and suppose that these intermediate elements supply α normal orders and β inversions in respect to k, and similarly α_1 normal orders and β_1 inversions in respect to p. We obviously have:

$$\alpha + \beta = \alpha_1 + \beta_1 = m. \tag{14}$$

As a result of the transposition, a normal order becomes an inversion and vice versa, or to put the matter more precisely, if element k was in normal order with regard to a certain inter-

mediate element before transposition, it becomes inverted after transposition and vice versa, whilst the same is true for element p. Thus the total number of inversions for elements k and p in regard to the intermediate elements was $\beta + \beta_1$ before transposition, and is $a+a_1$ after transposition, i.e. the change in the number of inversions is

$$\gamma = (a + a_1) - (\beta + \beta_1).$$

We can use (14) to re-write this as:

$$\gamma = (a + a_1) - (m - a + m - a_1) = 2\,(a + a_1 - m)\,,$$

whence it follows immediately that the number γ is even. We still have to take into account the change in the arrangement of elements k and p in regard to each other. If they were in normal order before transposition, they are afterwards inverted, and vice versa, i.e. the change in the number of inversions is unity here; hence the total change in the number of inversions due to transposition must be an odd number.

We notice some corollaries of the theorem.

Corollary I. If we write down all the $n!$ permutations and transpose two definite elements in each, say elements 1 and 3, all the even permutations become odd permutations and vice versa; whilst in general the total aggregate of $n!$ permutations is again obtained. It follows at once from this that *the numbers of even and odd permutations are the same.*

Corollary II. Every permutation can be obtained from the normal order by means of transpositions. It follows directly from the theorem that *even permutations are obtained by carrying out an even number of transpositions on the normal order, and odd permutations by carrying out an odd number of transpositions.*

Corollary III. The choice of normal order is entirely arbitrary. Any order other than (12) could have been taken as normal, in which case, of course, the definition of inversion would require a comparison with the new normal order. It may readily be seen that if we take any even permutation as normal instead of (12), even permutations still remain even, and similarly, odd permutations still remain odd. On the contrary, if we take any odd permutation as normal, even permutations become odd, and odd class permutations become even.

For instance, if we take 2, 1, 3 as the normal order in the six permutations of the elements 1, 2, 3, we have as even permutations:

$$2,\ 1,\ 3; \quad 1,\ 3,\ 2; \quad 3,\ 2,\ 1.$$

The second of these permutations contains two inversions: the 1 stands in front of the 2 and the 3 is in front of the 2, whereas in the normal order the 2 precedes the 1 and also precedes 3. The odd permutations are:

$$1,\ 2,\ 3; \quad 2,\ 3,\ 1; \quad 3,\ 1,\ 2.$$

We have one inversion with respect to the normal 2, 1, 3 in the first of these permutations, viz. 1 precedes 2.

On taking into account what has been said above, we can state the rule of signs in expression (8) as follows: *we write a plus sign in front of a product if the permutation of its second subscripts is even, and a minus sign if the permutation is odd, the order* $1, 2, \ldots, n$ *being taken as normal.*

We now elucidate one of the fundamental properties of determinants. We interchange the first and second columns in the array producing the determinant. The numbers written above as a_{ik} will still be denoted by the same letter with the same subscripts. Our interchange gives us, instead of array (6):

$$\left\| \begin{matrix} a_{12}, & a_{11}, & a_{13}, & \ldots, & a_{1n} \\ a_{22}, & a_{21}, & a_{23}, & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n2}, & a_{n1}, & a_{n3}, & \ldots, & a_{nn} \end{matrix} \right\|. \qquad (15)$$

We can now use the definition expressed by (8) to form the determinant corresponding to array (15). The columns in this array are enumerated by the following permutation: $2, 1, 3, \ldots, n$, and this must be taken as the normal order. It has been obtained from the previous normal order by means of one transposition, and therefore it was previously odd. Hence permutations that were previously odd become even with the new choice of normal order, and vice versa. It follows that the determinant corresponding to array (15) is the sum of the same terms as appear in (8) but, due to the change of the permutations of the second subscripts from odd to even or vice versa, all the terms now have the opposite signs, i.e. *the magnitude*

of a determinant changes sign on interchange of two columns. We have proved this property by interchanging the first and second columns. Exactly the same proof applies for the interchange of any two columns. We have, for instance:

$$\begin{vmatrix} 1, & 0, & 3 \\ 2, & 7, & 6 \\ 5, & 3, & 0 \end{vmatrix} = - \begin{vmatrix} 1, & 3, & 0 \\ 2, & 6, & 7 \\ 5, & 0, & 3 \end{vmatrix}.$$

The second determinant is obtained from the first by interchange of the second and third columns.

We consider a further property of determinants. A typical term of sum (8) is

$$(-1)^{[p_1, p_2, \ldots, p_n]} a_{1p_1} a_{2p_2} \ldots a_{np_n}. \tag{16$_1$}$$

We can bring the second subscripts into normal order by changing the order of the factors, but the first subscripts will now form some permutation $q_1, q_2, \ldots, q_n$, the expression being now written as

$$(-1)^{[p_1, p_2, \ldots, p_n]} a_{q_1 1} a_{q_2 2} \ldots a_{q_n n}. \tag{16$_2$}$$

The transition from (16_1) to (16_2) requires a certain number of transpositions of the factors. Each transposition implies a simultaneous transposition of both the first and the second subscripts. If the number of transpositions needed for passing from (16_1) to (16_2) is even, this means that the permutation $p_1, p_2, \ldots, p_n$ is even, since it becomes $1, 2, \ldots, n$ with the aid of an even number of transpositions. It can therefore be obtained similarly from the normal order with an even number of transpositions. But now the permutation $q_1, q_2, \ldots, q_n$ must likewise be even, since it is obtained simultaneously from the normal order with the aid of the same even number of transpositions. Similarly, if $p_1, p_2, \ldots, p_n$ is odd, $q_1, q_2, \ldots, q_n$ is too. It follows from this that $(-1)^{[p_1, p_2, \ldots, p_n]} = (-1)^{[q_1, q_2, \ldots, q_n]}$ and we can therefore write

$$(-1)^{[p_1, p_2, \ldots, p_n]} a_{1p_1} a_{2p_2} \ldots a_{np_n} = (-1)^{[q_1, q_2, \ldots, q_n]} a_{q_1 1} a_{q_2 2} \ldots a_{q_n n}.$$

Hence, if we compare corresponding terms in sums (8) and (10), it will be seen that the sums are precisely the same. The rows play the same role in sum (10) as the columns in sum (8). These remarks lead directly to the result that, *if all the rows and columns change*

places in an array without changing their order, the value of the determinant is unchanged.

For example, the following two third order determinants are equal:

$$\begin{vmatrix} 2, & 3, & 5 \\ 7, & 0, & 1 \\ 2, & 1, & 6 \end{vmatrix} = \begin{vmatrix} 2, & 7, & 2 \\ 3, & 0, & 1 \\ 5, & 1, & 6 \end{vmatrix}.$$

3. Fundamental properties of determinants. I. We first of all state the property just proved — *the value of a determinant is unchanged on replacing the rows by the columns.* Everything below that is proved for columns is likewise valid for rows, and vice versa.

II. We saw in the previous section that the interchange of two columns merely changes the sign of a determinant, the same being true for rows, i.e. *on interchange of two rows (columns) the determinant merely changes sign.*

III. If a determinant has two identical rows, on the one hand their interchange leads to the same determinant, whilst on the other hand, by what has been proved, the determinant changes sign. Thus, if we write the value of the determinant as Δ, we have $\Delta = -\Delta$, or $\Delta = 0$. In other words, *a determinant, with two identical rows (columns) is zero.*

IV. *A linear homogeneous function of the variables* $x_1, x_2, \ldots, x_n$ *is defined as a first degree polynomial in these variables with no constant term,* i.e. *it is an expression of the form*

$$\varphi(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \ldots + a_nx_n,$$

where the coefficients a_i are independent of the x_s. Such a function has two obvious properties:

$$\varphi(kx_1, kx_2, \ldots, kx_n) = k\varphi(x_1, x_2, \ldots, x_n),$$

$$\varphi(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) = \varphi(x_1, x_2, \ldots, x_n) + \varphi(y_1, y_2, \ldots, y_n).$$

The latter property remains valid for any number of added terms. On returning to formula (8), we see that each term of the sum contains one and only one element from each row as a factor. It follows from this that *a determinant is a linear homogeneous function of the elements of any given row (or of a given column).*

Consequently, *if all the elements of a row (column) contain a common factor, it can be taken outside the sign of the determinant.*

As indicated above, the value of the determinant corresponding to array (6) is generally denoted by

$$\begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} \end{vmatrix}$$

or, more briefly, by

$$|a_{ik}| \quad (i,\ k = 1,\ 2,\ \ldots,\ n).$$

The property just proved can be written in a particular case as say

$$\begin{vmatrix} ka_{11}, & ka_{12}, & ka_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{vmatrix} = k \begin{vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{vmatrix}.$$

The second property of linear homogeneous functions leads to the following property of determinants: *if the elements of a row (column) are the sums of a like number of terms, the determinant is equal to the sum of the determinants in which the elements of the row (column) in question are replaced by the individual terms.* We have, for example:

$$\begin{vmatrix} a, & b, & c+c' \\ d, & e, & f+f' \\ g, & h, & i+i' \end{vmatrix} = \begin{vmatrix} a, & b, & c \\ d, & e, & f \\ g, & h, & i \end{vmatrix} + \begin{vmatrix} a, & b, & c' \\ d, & e, & f' \\ g, & h, & i' \end{vmatrix}.$$

We note a further obvious consequence of linearity and homogeneity. *If all the elements of a given row (column) are zero, the determinant vanishes.*

V. If we strike out the ith row and kth column (intersecting in a_{ik}) from array (6), we are left with $(n-1)$ rows and $(n-1)$ columns. The corresponding $(n-1)$th order determinant is called the *minor* of the original nth order determinant, corresponding to the element a_{ik}. If we write this $\varDelta_{ik}$, the product

$$A_{ik} = (-1)^{i+k}\varDelta_{ik} \tag{17}$$

is called the *cofactor* of the element a_{ik}. We now show that these cofactors are the coefficients of the linear homogeneous function referred to in an earlier property, i.e. we have for any ith row:

$$\varDelta = A_{i1}a_{i1} + A_{i2}a_{i2} + \ldots + A_{in}a_{in} \quad (i = 1,\ 2,\ \ldots,\ n) \tag{18}$$

and for any kth column:

$$\varDelta = A_{1k}a_{1k} + A_{2k}a_{2k} + \ldots + A_{nk}a_{nk} \quad (k = 1,\ 2,\ \ldots,\ n), \tag{19}$$

where Δ is the value of the determinant. We have to show, in other words, that if we collect all the terms in (8) containing a given element a_{ik}, the coefficient of the element will be its cofactor A_{ik}, as defined by (17). We write this coefficient to start with as B_{ik} and observe that it consists of the sum of the products of $(n - 1)$ elements, elements of the ith row and kth column being no longer included in these products.

We first take the case $i = k = 1$ and write out the terms of sum (8) containing a_{11}:

$$a_{11} \sum_{(p_2, \ldots, p_n)} (-1)^{[1, p_2, \ldots, p_n]} a_{2p_2} \ldots a_{np_n}.$$

The summation must extend over all the possible permutations $p_2, p_3, \ldots, p_n$ of the numbers $2, 3, \ldots, n$. The first element unity in the full permutation $1, p_2, \ldots, p_n$ is in the normal order as regards the remaining elements, so that we have for the number of inversion:

$$[1, p_2, \ldots, p_n] = [p_2, \ldots, p_n],$$

the permutation in which the numbers appear in increasing order being taken as normal in both cases. We thus have the following expression for the coefficient of a_{11}:

$$B_{11} = \sum_{(p_2, p_3, \ldots, p_n)} (-1)^{[p_2, \ldots, p_n]} a_{2p_2} \ldots a_{np_n}.$$

This sum comes within the definition of determinant, except that, by comparison with the original determinant, the first row and first column are missing. Hence it is clear that

$$B_{11} = \Delta_{11} = (-1)^{1+1} \Delta_{11} = A_{11},$$

i.e. our statement is proved for $i = k = 1$. We turn to the case of any i and k. We interchange the ith row in turn with higher rows until it arrives at the position of the first row. This requires $(i - 1)$ interchanges of rows. Similarly, the kth column is brought to the position of the first column by successive interchanges. These interchanges move the element a_{ik} upwards and to the left to the position of a_{11}. The row characterized by i and the column characterized by k appear in the first position, whilst the order of the remaining rows and columns remains unchanged. The result obtained above shows that the coefficient of a_{ik} after these interchanges is equal to Δ_{ik}. But we have employed $(i - 1) + (k - 1)$ interchanges of rows and columns in pairs, and each such interchange adds a factor of (-1)

to the determinant, i.e. we have added altogether the factor

$$(-1)^{(i-1)+(k-1)} = (-1)^{i+k},$$

and the final expression for the coefficient B_{ik} is therefore:

$$B_{ik} = \frac{\Delta_{ik}}{(-1)^{i+k}} = (-1)^{i+k}\Delta_{ik} = A_{ik}$$

which is what we wished to prove. Thus we have proved formulae (18) and (19). If we successively replace the elements of the ith row in the determinant Δ by the numbers $c_1, c_2, \ldots, c_n$, whilst the remaining rows are unchanged, the factors A_{is} in (18) will be unchanged, and the value of the new determinant will be

$$\Delta' = A_{i1}c_1 + A_{i2}c_2 + \ldots + A_{in}c_n. \tag{20}$$

In particular, if we take $c_1, c_2, \ldots, c_n$ equal to the elements a_{j1}, $a_{j2}, \ldots, a_{jn}$ of the jth row, where $j \neq i$, the determinant Δ' will have two identical rows, the ith and the jth, so that it vanishes: $\Delta' = 0$, i.e.

$$A_{i1}a_{j1} + A_{i2}a_{j2} + \ldots + A_{in}a_{jn} = 0 \qquad (i \neq j). \tag{21$_1$}$$

Similarly for columns:

$$A_{1k}a_{1l} + A_{2k}a_{2l} + \ldots + A_{nk}a_{nl} = 0 \qquad (k \neq l). \tag{21$_2$}$$

Expressions (19) and (21) lead us to a property of determinants that is important later.

If we multiply the elements of a row (column) by their cofactors then add, we get the value of the determinant. On the other hand, if we multiply the elements of a row (column) by the cofactors of the corresponding elements of another row (column) then add, the sum is zero.

VI. We add the elements of the second row, multiplied by a factor p, to the elements of the first row of the determinant Δ. The elements of the first row become

$$a_{1s} + pa_{2s} \quad (s = 1, 2, \ldots, n),$$

and by virtue of property IV, the new determinant is the sum of two determinants: the original determinant and a second determinant in which the first row consists of the elements

$$pa_{2s} \quad (s = 1, 2, \ldots, n),$$

whilst the remaining rows are the same as in Δ. On taking p out of the first row, we get identical first and second rows, and the second determinant, therefore, vanishes, i.e. in general, *the value of a deter-*

minant is unchanged if the elements of one row (column), multiplied by a constant factor, are added to the corresponding elements of another row (column).

We now introduce some notation for future use. Given the square array of numbers (6), let l be a positive integer not greater than n. We shall denote the determinant of order l, consisting of the rows of (6) numbered $p_1, p_2, \ldots, p_l$ and the columns of (6) numbered $q_1, q_2, \ldots, q_l$, as follows:

$$A\begin{pmatrix} p_1, p_2, \ldots, p_l \\ q_1, q_2 \ldots, q_l \end{pmatrix} = \begin{vmatrix} a_{p_1q_1}, & a_{p_1q_2}, & \ldots, & a_{p_1q_l} \\ a_{p_2q_1}, & a_{p_2q_2}, & \ldots, & a_{p_2q_l} \\ \cdot & \cdot & \cdot & \cdot \\ a_{p_lq_1}, & a_{p_lq_2}, & \ldots, & a_{p_lq_l} \end{vmatrix}. \tag{22}$$

With this, a number a itself is generally referred to as the first order determinant corresponding to a, i.e. $A\binom{p}{q} = a_{pq}$. The sequences of positive integers $p_1, p_2, \ldots, p_l$ and $q_1, q_2, \ldots, q_l$ need not necessarily be arranged in order of increasing p_s and q_s. If the numbers are in fact in increasing order in both sequences, determinant (22) is called a *minor* of order l of determinant (8). The determinant (22) is obtained from (8) by striking out $(n - l)$ rows and $(n - l)$ columns. Let the rows and columns struck out be characterized by the numbers r_1, $r_2, \ldots, r_{n-l}$ and $s_1, s_2, \ldots, s_{n-l}$, arranged in increasing order. The minor

$$A\begin{pmatrix} r_1, r_2, \ldots, r_{n-l} \\ s_1, s_2, \ldots, s_{n-l} \end{pmatrix}$$

is known as the complementary minor to (22), whilst the expression

$$(-1)^{p_1+p_2+\ldots+p_l+q_1+q_2+\ldots+q_l} A\begin{pmatrix} r_1, r_2, \ldots, r_{n-l} \\ s_1, s_2, \ldots, s_{n-l} \end{pmatrix} \tag{22_1}$$

is known as the cofactor to minor (22). For a single element a_{ik}, this definition of cofactor is the same as the previous definition (17).

We write the cofactor (22_1) as

$$A'\begin{pmatrix} p_1, & p_2, & \ldots, & p_l \\ q_1, & q_2, & \ldots, & q_l \end{pmatrix}.$$

It is fully defined on specifying determinant (22), i.e. on specifying the numerical sequences $p_1, p_2, \ldots, p_l$ of its rows and $q_1, q_2, \ldots, q_l$ of its columns.

Let us fix the numbers of the rows. The value of the determinant Δ is evidently a homogeneous polynomial of degree l in the elements of these rows, and it can be shown to be given by the expression (Laplace's theorem):

$$\Delta = \sum_{q_1 < q_2 \ldots < q_l} A \begin{pmatrix} p_1, & p_2, & \ldots, & p_l \\ q_1, & q_2, & \ldots, & q_l \end{pmatrix} A' \begin{pmatrix} p_1, & p_2, & \ldots, & p_l \\ q_1, & q_2, & \ldots, & q_l \end{pmatrix}, \tag{23}$$

where the summation is carried out over all the possible increasing sequences of $q_1, q_2, \ldots, q_l$ taken from the sequence $1, 2, \ldots, n$. The number of terms in sum (23) is equal to the number of combinations of l from n elements:

$$C_n^l = \frac{n(n-1)\ldots(n-l+1)}{l!},$$

since the q_s are only taken in increasing order when forming the sum. For $l=1$, we have $A\binom{p_1}{q_1} = a_{p_1 q_1}$, and (23) becomes (18) with $i = p_1$. It is easy to derive an expression analogous to (23) for the expansion of Δ in the elements of any selected columns. No use will be made of (23) and the proof is omitted.

4. Evaluation of determinants. The evaluation of a second order determinant is extremely simple. By (11), we merely write down the array

$$\left\| \begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix} \right\|$$

and take the product of the elements on the full diagonal with its own sign, and on the dotted diagonal with the reverse sign.

We turn to third order determinants. We wrote down the expanded form in (3). It may easily be verified that this can be obtained as follows: we write out the array giving the determinant then write out the first and second rows again underneath. This gives us an array with six diagonals, with three elements on each. We take the products of the elements on full line diagonals without change,

$$\left\| \begin{matrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{matrix} \right\|$$

and those of the elements on dotted diagonals with the minus sign.

The sum of the six products gives the determinant (Sarrus' rule).

There is no generalization of this rule for higher order determinants and we need a different procedure to shorten the working. For instance, property VI of the previous section can often be used with advantage. We shall illustrate this with an example. We take the fourth order determinant

$$\Delta = \begin{vmatrix} 3, & 5, & 1, & 0 \\ 2, & 1, & 4, & 5 \\ 1, & 7, & 4, & 2 \\ -3, & 5, & 1, & 1 \end{vmatrix}.$$

We multiply the third row by (-2) and add it to the second; further, we multiply the same row by 3 and add it to the fourth whilst subtracting from the first. By the property mentioned, we arrive at a determinant of the same value as that written above, but now having three zeros in the first column:

$$\Delta = \begin{vmatrix} 0, & -16, & -11, & -6 \\ 0, & -13, & -4, & 1 \\ 1, & 7, & 4, & 2 \\ 0, & 26, & 13, & 7 \end{vmatrix}.$$

This gives us, on expanding by elements of the first column in accordance with equation (19):

$$\Delta = \begin{vmatrix} -16, & -11, & -6 \\ -13, & -4, & 1 \\ 26, & 13, & 7 \end{vmatrix}.$$

We multiply the third column by 4 and add it to the second, then multiply the same column by 13 and add it to the first. We thus get:

$$\Delta = \begin{vmatrix} -94, & -35, & -6 \\ 0, & 0, & 1 \\ 117, & 41, & 7 \end{vmatrix} = -\begin{vmatrix} -94, & -35 \\ 117, & 41 \end{vmatrix} = 94\times 41 - 35\times 117 = -241.$$

5. Examples. 1. Let it be required to find the volume of a parallelepiped whose sides are the vectors **A**, **B**, **C**, having the same vertex as origin. We know from [II, 105] that the required volume is given by the scalar product of **A** and the vector product $(\mathbf{B}\times\mathbf{C})$:

$$V = \mathbf{A}\,(\mathbf{B}\times\mathbf{C}). \tag{24}$$

The volume is obtained with the plus sign if **A**, **B**, **C** have the same orientation as the coordinate axes, and with the minus sign if the orientation is different. The components of the vector product are

$$B_yC_z - B_zC_y, \quad B_zC_x - B_xC_z, \quad B_xC_y - B_yC_x,$$

and the scalar product of (24) is therefore

$$A_x(B_yC_z - B_zC_y) + A_y(B_zC_x - B_xC_z) + A_z(B_xC_y - B_yC_x).$$

This sum may easily be seen to represent a third order determinant, i.e.

$$V = \begin{vmatrix} A_x, & B_x, & C_x \\ A_y, & B_y, & C_y \\ A_z, & B_z, & C_z \end{vmatrix}. \tag{25}$$

The vanishing of this determinant shows us that the volume is zero, in other words, that the three vectors are coplanar, i.e. they lie in one plane. If we interchange two rows (columns) in the determinant, say the first and second, this means that the order of vectors **A**, **B**, **C** is changed to **B**, **A**, **C**; if the vectors had the same orientation as the axes in the previous sequence, they now have a different orientation, and vice versa. The value of the determinant correspondingly changes sign.

Similarly, if we take two vectors (A_x, A_y) and (B_x, B_y) in the xy plane, the area of the parallelogram formed by them is equal to the second order determinant

$$P = \begin{vmatrix} A_x, & B_x \\ A_y, & B_y \end{vmatrix}.$$

We now consider a triangle with vertices

$$M_1(x_1, y_1), \quad M_2(x_2, y_2), \quad M_3(x_3, y_3).$$

We take the vector $\mathbf{A} = \overline{M_1M_2}$ and $\mathbf{B} = \overline{M_1M_3}$, with components:

$$\overline{M_1M_2}(x_2 - x_1, \; y_2 - y_1), \; \overline{M_1M_3}(x_3 - x_1, \; y_3 - y_1),$$

whilst the area of the triangle can be written as

$$P = \frac{1}{2}\begin{vmatrix} x_2 - x_1, & x_3 - x_1 \\ y_2 - y_1, & y_3 - y_1 \end{vmatrix}.$$

It may readily be shown that the second order determinant can be replaced by a third order determinant so that the above expression becomes

$$P = \frac{1}{2}\begin{vmatrix} x_1, & x_2, & x_3 \\ y_1, & y_2, & y_3 \\ 1, & 1, & 1 \end{vmatrix}.$$

The vanishing of this determinant gives the condition for the three points M_1, M_2, M_3 to be collinear. In other words, the equation of the straight line passing through two given points (x_1, y_1) and (x_2, y_2) can be written as

$$\begin{vmatrix} x, & x_1, & x_2 \\ y, & y_1, & y_2 \\ 1, & 1, & 1 \end{vmatrix} = 0.$$

II. The equations of some loci may easily be found by using determinants. Suppose, for instance, we are seeking the equation of the circle passing through three given points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$. The equation may be readily seen to be obtainable with the aid of a fourth order determinant, as follows:

$$\begin{vmatrix} x^2 + y^2, & x_1^2 + y_1^2, & x_2^2 + y_2^2, & x_3^2 + y_3^2 \\ x, & x_1, & x_2, & x_3 \\ y, & y_1, & y_2, & y_3 \\ 1, & 1, & 1, & 1 \end{vmatrix} = 0. \qquad (26)$$

On expanding by the first column, this equation is seen to be in fact of the second degree, with the same coefficient for x^2 as y^2 and with the term in xy missing, i.e. (26) is the equation of a circle. Finally, if we substitute $x = x_k$ and $y = y_k$ in the equation $(k = 1, 2, 3)$, the first column becomes identical with one of the others and the equation is satisfied, i.e. the circle actually passes through the three given points. It should be noted that, if the three given points are collinear, the coefficient of $(x^2 + y^2)$ in equation (26) vanishes, so that the equation corresponds to a straight line and not a circle.

Similarly, with axes OX, OY, OZ in space, the equation of the plane passing through three given points (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) can be written as a fourth order determinant:

$$\begin{vmatrix} x, & x_1, & x_2 & x_3 \\ y, & y_1, & y_2 & y_3 \\ z, & z_1, & z_2, & z_3 \\ 1, & 1, & 1, & 1 \end{vmatrix} = 0. \qquad (27)$$

If the three given points are collinear, equation (27) reduces to the identity $0 = 0$.

III. We consider the determinant D_n of order n, each row of which consists of powers of a certain number, starting with the $(n-1)$th power and down to and including zero:

$$D_n = \begin{vmatrix} x_1^{n-1}, x_1^{n-2}, \ldots, x_1, 1 \\ x_2^{n-1}, x_2^{n-2}, \ldots, x_2, 1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ x_n^{n-1}, x_n^{n-2} \ldots, x_n, 1 \end{vmatrix}. \tag{28}$$

We have with $n=1$ and $n=2$:

$$D_1 = 1;\ D_2 = x_1 - x_2.$$

To expand the determinant D_3, we replace the number x_1 in its first row by the letter x. We get the determinant:

$$D_3(x) = \begin{vmatrix} x^2, & x, & 1 \\ x_2^2, & x_2, & 1 \\ x_3^2, & x_3, & 1 \end{vmatrix}.$$

On expanding by the first column, we see that $D_3(x)$ is a second degree polynomial in x. If we substitute $x = x_2$ or $x = x_3$ in the determinant, the first row becomes the same as the second or third and the determinant has zero value, i.e. the quadratic form $D_3(x)$ has roots x_2 and x_3 and may be written as

$$D_3(x) = A_3(x - x_2)(x - x_3),$$

where A_3 is the coefficient of x^2, i.e. the cofactor of the element x^2 appearing at the top left-hand corner of $D_3(x)$. It follows from this that

$$A_3 = \begin{vmatrix} x_2, & 1 \\ x_3, & 1 \end{vmatrix},$$

i.e. A_3 is the determinant D_2, consisting of the numbers x_2 and x_3. Finally:

$$D_3(x) = (x_2 - x_3)(x - x_2)(x - x_3).$$

On substituting $x = x_1$, we obtain an expression for D_3 as the product of three factors:

$$D_3 = \begin{matrix} (x_1 - x_2)(x_1 - x_3) \\ \qquad\qquad (x_2 - x_3). \end{matrix}$$

Having found D_3, we can find an expression for D_4 in precisely the same way. It is the product of six factors:

$$D_4 = \begin{array}{r} (x_1 - x_2)(x_1 - x_3)(x_1 - x_4) \\ (x_2 - x_3)(x_2 - x_4) \\ (x_3 - x_4) \end{array}$$

Similarly, for any n, we find the following expression for D_n, which is generally known as Vandermonde's determinant:

$$D_n = \begin{array}{r} (x_1 - x_2)(x_1 - x_3) \ldots (x_1 - x_n) \\ (x_2 - x_3) \ldots (x_2 - x_n). \\ \cdots\cdots\cdots\cdots\cdots \\ (x_{n-1} - x_n) \end{array} \tag{29}$$

This expression has an interesting connection with the basic definition of determinant. Any nth order determinant can be written:

$$\begin{vmatrix} x_{1n}, & x_{1,n-1}, & \ldots, & x_{11} \\ x_{2n}, & x_{2,n-1}, & \ldots, & x_{21} \\ \cdot & \cdot & \cdot & \cdot \\ x_{nn}, & x_{n,n-1}, & \ldots, & x_{n1} \end{vmatrix}. \tag{30}$$

We carry out the purely formal substitution of x_i^{k-1} for each element x_{ik}. As a result of this, determinant (30) clearly becomes Vandermonde's determinant (28). An immediate consequence is the following rule for finding the sum giving the value of (30): we remove the brackets in expression (29) and replace x_k^{s-1} in each of the resultant terms by x_{ks}; if a power of x_k is missing in a product term, we add the factor x_k^0, which after substitution becomes x_{k1}. It may be remarked that this last rule can be taken as the definition of a determinant.

IV. We consider an expression which will concern us later on:

$$\Delta(x) = \begin{vmatrix} a_{11} + x, & a_{12}, & a_{13}, & \ldots, & a_{1n} \\ a_{21}, & a_{22} + x, & a_{23}, & \ldots, & a_{2n} \\ a_{31}, & a_{32}, & a_{33} + x, & \ldots, & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & a_{n3}, & \ldots, & a_{nn} + x \end{vmatrix} \tag{31}$$

and expand it in powers of x; for this, we first re-write it as follows:

$$\Delta(x) = \begin{vmatrix} a_{11}+x, & a_{12}+0, & a_{13}+0, & \ldots, & a_{1n}+0 \\ a_{21}+0, & a_{22}+x, & a_{23}+0, & \ldots, & a_{2n}+0 \\ a_{31}+0, & a_{32}+0, & a_{33}+x, & \ldots, & a_{3n}+0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1}+0, & a_{n2}+0, & a_{n3}+0, & \ldots, & a_{nn}+x \end{vmatrix}. \tag{32}$$

Each column of this determinant is the sum of two terms and we can re-write it by means of repeated application of property IV above as the sum of 2^n determinants, the columns of which contain no sums. If we strike out the second term in all the columns of (32), we get a term which does not include x, i.e. the constant term in the expansion of $\Delta(x)$:

$$\Delta = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} \end{vmatrix}. \tag{33}$$

On the other hand, if we strike out the first terms in all the columns, we get the leading term of polynomial $\Delta(x)$:

$$\begin{vmatrix} x, & 0, & 0, & \ldots, & 0 \\ 0, & x, & 0, & \ldots, & 0 \\ 0, & 0, & x, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \ldots, & x \end{vmatrix} = x^n.$$

We now consider the middle terms of the polynomial. Suppose we retain the second term in the $p_1, p_2, \ldots, p_s$th columns, and retain the first term in the remaining columns. Each p_kth column ($k = 1, 2, \ldots, s$) will now consist entirely of zeros except for the single element x on the principal diagonal, i.e. on the intersections of rows and columns characterized by the same number. On successively expanding the present determinant by the $p_1, p_2, \ldots, p_s$th columns, we get the factor x^s from these columns and have to strike out the $p_1, p_2, \ldots, p_s$th rows and columns. The cofactor of the powers of x after each striking out is precisely equal to the minor since the row and column struck out are both characterized by the same number. It follows that, for any choice of columns p_k ($k = 1, 2, \ldots, s$), our determinant will contain x^s with a coefficient equal to the determinant of order

$(n - s)$ obtained from the original determinant (33) by striking out the rows and columns whose intersections consist of the elements $a_{p_1p_1}$, $a_{p_2p_2}, \ldots, a_{p_sp_s}$ of the principal diagonal. We write this $(n - s)$th order determinant symbolically as $\Delta_{\substack{p_1p_2 \ldots p_s \\ p_1p_2 \ldots p_s}}$.

This is usually called the principal minor of order $(n - s)$ *of the determinant* Δ. Different choices of $p_1, p_2, \ldots, p_s$ lead us eventually to the final coefficient of x^s in the expression for $\Delta(x)$ as the sum of all the possible principal minors of order $(n - s)$, i.e.

$$\Delta(x) = x^n + S_1 x^{n-1} + S_2 x^{n-2} + \ldots + S_{n-1} x + S_n,$$

where S_k is the sum of all the kth order principal minors of Δ. In particular, $S_n = \Delta$. The expression for the coefficient is explicitly

$$S_k = \sum_{p_1 < p_2 < \ldots < p_{n-k}}^{(1, 2, \ldots, n)} \Delta_{\substack{p_1p_2 \ldots p_{n-k} \\ p_1p_2 \ldots p_{n-k}}} = \sum_{q_1 < q_2 < \ldots < q_k} \begin{vmatrix} a_{q_1q_1}, & a_{q_1q_2}, & \ldots, & a_{q_1q_k} \\ a_{q_2q_1}, & a_{q_2q_2}, & \ldots, & a_{q_2q_k} \\ \cdot & \cdot & \cdot & \cdot \\ a_{q_kq_1}, & a_{q_kq_2}, & \ldots, & a_{q_kq_k} \end{vmatrix}. \tag{34}$$

Here the summation extends over all the possible combinations of the k numbers $q_1, q_2, \ldots, q_k$, taken in increasing order from among the numbers $1, 2, \ldots, n$. If the summation in (34) were simply over each subscript q_j for all values from 1 to n, the integers would appear in the permutation $q_1, q_2, \ldots, q_k$ in all possible orders and not solely in increasing order. To be precise, every increasing sequence in the summation over all q_j from 1 to n would have its place taken by $k!$ permutations in all. We now observe that the magnitude of the determinant appearing in (34) is unchanged on interchange of any two numbers q_i and q_j. Suppose, for instance, that q_1 and q_2 are interchanged, then the first and second rows and columns are interchanged in the determinant which has no effect on its value. It follows from these remarks that, if the summation in (34) is simply over each of the q_j from 1 to n, each term in sum (34) will be repeated $k!$ times, so that we can write the coefficient S_k in the alternative form:

$$S_k = \frac{1}{k!} \sum_{q_1=1}^{n} \sum_{q_2=1}^{n} \ldots \sum_{q_k=1}^{n} \Delta \begin{pmatrix} q_1, & q_2, & \ldots, & q_k \\ q_1, & q_2, & \ldots, & q_k \end{pmatrix}. \tag{35}$$

6. Multiplication of determinants. We derive a formula in this article for the product of two determinants of the same order.

Let us be given two nth order determinants:

$$\Delta = |a_{ik}|_1^n \tag{36_1}$$

and

$$\Delta_1 = |b_{ik}|_1^n. \tag{36_2}$$

We form a new determinant, the elements of which are given by

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{sk} \qquad (i,\ k = 1, 2, \ldots,\ n) \tag{37}$$

and we show that this determinant is equal to the product of determinants (36_1) and (36_2). We start with the case $n = 2$. On taking into account (37) and expanding the determinant in accordance with property IV of [3], we get:

$$\begin{vmatrix} c_{11}, c_{12} \\ c_{21}, c_{22} \end{vmatrix} = \begin{vmatrix} a_{11}b_{11} + a_{12}b_{21}, a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21}, a_{21}b_{12} + a_{22}b_{22} \end{vmatrix} = \begin{vmatrix} a_{11}b_{11}, & a_{11}b_{12} \\ a_{21}b_{11}, & a_{21}b_{12} \end{vmatrix} +$$
$$+ \begin{vmatrix} a_{11}b_{11}, & a_{12}b_{22} \\ a_{21}b_{11}, & a_{22}b_{22} \end{vmatrix} + \begin{vmatrix} a_{12}b_{21}, & a_{11}b_{12} \\ a_{22}b_{21}, & a_{21}b_{12} \end{vmatrix} + \begin{vmatrix} a_{12}b_{21}, & a_{12}b_{22} \\ a_{22}b_{21}, & a_{22}b_{22} \end{vmatrix}.$$

On taking outside the common factors of the same columns, the first and fourth terms on the right-hand side yield determinants with identical columns and these vanish. On interchanging columns in one of the determinants that remain, we find that

$$\begin{vmatrix} c_{11}, c_{12} \\ c_{21}, c_{22} \end{vmatrix} = b_{11}b_{22} \begin{vmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{vmatrix} + b_{12}b_{21} \begin{vmatrix} a_{12}, a_{11} \\ a_{22}, a_{21} \end{vmatrix} =$$
$$= b_{11}b_{22} \begin{vmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{vmatrix} - b_{12}b_{21} \begin{vmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{vmatrix} = \begin{vmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{vmatrix} (b_{11}b_{22} - b_{12}b_{21}) =$$
$$= \begin{vmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{vmatrix} \cdot \begin{vmatrix} b_{11}, b_{12} \\ b_{21}, b_{22} \end{vmatrix}$$

which is what we had to prove.

In the general case of order n, we have after applying property IV of [3]:

$$|c_{ik}|_1^n = \sum_{s_1, s_2, \ldots, s_n} \begin{vmatrix} a_{1s_1}b_{s_1 1}, a_{1s_2}b_{s_2 2}, \ldots, a_{1s_n}b_{s_n n} \\ a_{2s_1}b_{s_1 1}, a_{2s_2}b_{s_2 2}, \ldots, a_{2s_n}b_{s_n n} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{ns_1}b_{s_1 1}, a_{ns_2}b_{s_2 2}, \ldots, a_{ns_n}b_{s_n n} \end{vmatrix}. \tag{38}$$

where the variables of summation $s_1, s_2, \ldots, s_n$ take the integral values $1, 2, \ldots, n$. The terms of this sum can be written:

$$b_{s_1 1} b_{s_2 2} \ldots b_{s_n n} \begin{vmatrix} a_{1s_1}, & a_{1s_2}, & \ldots, & a_{1s_n} \\ a_{2s_1}, & a_{2s_2}, & \ldots, & a_{2s_n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{ns_1}, & a_{ns_2}, & \ldots, & a_{ns_n} \end{vmatrix}. \tag{39}$$

If any of the numbers $s_1, s_2, \ldots, s_n$ are the same, there will be equal columns in the above determinant and it vanishes. We can thus confine ourselves to the terms for which none of the s_k are the same, so that the sequence $s_1, s_2, \ldots, s_n$ represents a permutation of $1, 2, \ldots, n$. Twice multiplying (39) by $(-1)^{[s_1, s_2, \ldots, s_n]}$ evidently leaves the expression unchanged, so that we can write it as the product of two factors:

$$(-1)^{[s_1, s_2, \ldots s_n]} \begin{vmatrix} a_{1s_1}, & a_{1s_2}, & \ldots, & a_{1s_n} \\ a_{2s_1}, & a_{2s_2}, & \ldots, & a_{2s_n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{ns_1}, & a_{ns_2}, & \ldots, & a_{ns_n} \end{vmatrix} \cdot (-1)^{[s_1, s_2, \ldots, s_n]} b_{s_1 1} b_{s_2 2} \ldots b_{s_n n}.$$

We transpose in the first factor so that the sequence $s_1, s_2, \ldots, s_n$ becomes $1, 2, \ldots, n$. Each transposition of $(-1)^{[s_1, s_2, \ldots, s_n]}$ simply changes the sign of the determinant, whilst the factor as a whole remains unchanged. Hence we can write (39) as

$$\begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} \end{vmatrix} \cdot (-1)^{[s_1, s_2, \ldots, s_n]} b_{s_1 1} b_{s_2 2} \ldots b_{s_n n},$$

and we now obtain, on returning to sum (38):

$$|c_{ik}|_1^n = \Delta \sum_{(s_1, s_2, \ldots, s_n)} (-1)^{[s_1, s_2, \ldots, s_n]} b_{s_1 1} b_{s_2 2} \ldots b_{s_n n},$$

where the summation extends over all the permutations $s_1 s_2, \ldots, s_n$ of the numbers $1, 2, \ldots, n$. This latter sum is the determinant Δ_1, i.e. $|c_{ik}|_1^n = \Delta\Delta_1$ which is what we had to show. Equation (37) amounts to the following: the elements of the ith row of determinant Δ are multiplied by the corresponding elements of the kth column of the second determinant then the products added. We know that the rows can be replaced by the columns in a determinant without changing its value. The above rule for multiplying rows by columns can there-

fore be replaced by three alternative rules, for multiplying columns by columns, columns by rows, and rows by rows.

We can finally state the theorem as: *let*

$$|a_{ik}| \quad \text{and} \quad |b_{ik}|$$

be two determinants of any order n.

We form a new determinant

$$|c_{ik}|,$$

whose elements are given by one of the following expressions:

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{sk}, \tag{40$_1$}$$

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{ks} \tag{40$_2$}$$

$$c_{ik} = \sum_{s=1}^{n} a_{si} b_{sk}, \tag{40$_3$}$$

$$c_{ik} = \sum_{s=1}^{n} a_{si} b_{ks} \quad (i, k = 1, 2, \ldots, n). \tag{40$_4$}$$

The value of the determinant $|c_{ik}|$ *is now equal to the product of* $|a_{ik}|$ *and* $|b_{ik}|$.

Example. We consider, along with the original determinant

$$\Delta = |a_{ik}|$$

the determinant consisting of the cofactors of its elements

$$|A_{ik}|.$$

We shall express the product $|a_{ik}| \cdot |A_{ik}|$ as a further determinant, by multiplying rows by rows, in accordance with the above theorem. The new determinant has the following elements:

$$c_{ik} = \sum_{s=1}^{n} a_{is} A_{ks}.$$

We obtain from property V of determinants:

$$c_{ik} = 0 \text{ for } i \neq k; \quad c_{ii} = \Delta,$$

i.e.

$$|a_{ik}|\cdot|A_{ik}| = \begin{vmatrix} \Delta, & 0, & 0, & \dots, & 0 \\ 0, & \Delta, & 0, & \dots, & 0 \\ 0, & 0, & \Delta, & \dots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \dots, & \Delta \end{vmatrix}$$

or, as may readily be seen:

$$|a_{ik}|_1^n |A_{ik}|_1^n = \Delta^n, \qquad \text{i.e.} \qquad \Delta\,|A_{ik}|_1^n = \Delta^n.$$

We have on dividing through by Δ, assumed non-zero:

$$|A_{ik}|_1^n = \Delta^{n-1}. \tag{41}$$

If the elements $a_{ik} = a_{ik}^{(0)}$ are such that Δ vanishes, we can find elements a_{ik} as near as we like to $a_{ik}^{(0)}$ such that Δ differs from zero. Equation (41) is valid for these a_{ik}, and on passing to the limit with $a_{ik} \to a_{ik}^{(0)}$, we see that the equation remains valid for $a_{ik} = a_{ik}^{(0)}$, i.e. for $\Delta = 0$. If Δ and A_{ik} are written in terms of the elements a_{ik}, (41) represents an identity with respect to the a_{ik}.

7. Rectangular arrays. We shall encounter later on arrays in which the number of rows can differ from the number of columns. This more general type of array is exemplified by

$$\begin{Vmatrix} a_{11}, & a_{12}, & \dots, & a_{1n} \\ a_{21}, & a_{22}, & \dots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1}, & a_{m2}, & \dots, & a_{mn} \end{Vmatrix}. \tag{42}$$

It contains m rows and n columns, where the m and n can be the same or different. On striking out rows and columns so that the number of each is the same, we can form determinants from the remainder. We say that these determinants enter into the constitution of array (42). The highest order that they can have is evidently equal to the lesser of the two numbers m and n, whilst the least order is unity, the first order determinants being in fact the actual elements of array (42). Suppose that all the determinants of a certain order l appearing in the array are zero. It may readily be seen that all the determinants of order $(l + 1)$ in the array are likewise zero. In fact, each determinant of order $(l + 1)$ can be expressed as the sum of the products of the elements of a given row with the cofactors

of these elements. But the latter, except for sign, coincide with determinants of order l of the array, and are therefore all zero. Since all the determinants of order $(l+1)$ are zero, it follows as above that all the determinants of order $(l+2)$ likewise vanish, and so on. Thus, if all the determinants of a given order appearing in array (42) vanish, all the higher order determinants of the array likewise vanish.

This brings us to an important concept regarding array (42), the array being more commonly known as a matrix. *The rank of matrix* (42) *is defined as the highest order non-vanishing determinant of the matrix*, i.e. *if the rank is k, there is at least one non-vanishing determinant of order k in the matrix, whereas all the determinants of order $(k+1)$ vanish.*

Let us consider, along with matrix (42), the array

$$\begin{Vmatrix} b_{11}, & b_{12}, & \ldots, & b_{1m} \\ b_{21}, & b_{22}, & \ldots, & b_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ b_{n1}, & b_{n2}, & \ldots, & b_{nm} \end{Vmatrix}, \tag{43}$$

consisting of n rows and m columns. We form the m^2 numbers

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{sk} \quad (i,\ k = 1,\ 2,\ \ldots,\ m). \tag{44}$$

The square array made up of the c_{ik} is usually known as the product of rectangular arrays (42) and (43).

We prove a generalization of the theorem for multiplication of determinants.

THEOREM. *If $m \leqslant n$, we have*

$$|c_{ik}|_1^m = \sum_{r_1 < r_2 < \ldots < r_m} A\begin{pmatrix} 1, & 2, & \ldots, & m \\ r_1, & r_2, & \ldots, & r_m \end{pmatrix} B\begin{pmatrix} r_1, & r_2, & \ldots, & r_m \\ 1, & 2, & \ldots, & m \end{pmatrix}, \tag{45}$$

where the summation extends over all the r_k of the sequence $1, 2, \ldots, n$, *satisfying the inequalities indicated. If $m > n$, the determinant $|c_{ik}|_1^m$ vanishes.*

The meaning of the symbols

$$A\begin{pmatrix} 1, & 2, & \ldots, & m \\ r_1 & r_2, & \ldots, & r_m \end{pmatrix} \quad \text{and} \quad B\begin{pmatrix} r_1, & r_2, & \ldots, & r_m \\ 1, & 2, & \ldots, & m \end{pmatrix}$$

is given in [3]. The second denotes the determinant formed from the elements of the $r_1, r_2, \ldots, r_m$th row and first, second, $\ldots$, mth

columns of array (43). For $m=n$, the sum in (45) reduces to the single term corresponding to $r_1 = 1, r_2 = 2, \ldots, r_m = m$, and (45) expresses the theorem for multiplication of determinants.

We take the case $m < n$. The proof of (45) is analogous to that for the multiplication of determinants; we have as in the latter case:

$$|c_{ik}|_1^m = \sum_{s_1, \ldots, s_m} A\begin{pmatrix}1, & 2, & \ldots, & m\\ s_1, & s_2, & \ldots, & s_m\end{pmatrix} b_{s_1 1} b_{s_2 2} \ldots b_{s_m m}, \tag{46}$$

where each of the s_k can take the values $1, 2, \ldots, n$, and where terms can be neglected for which some of the s_k are equal, since such terms are zero. We take a definite sequence of numbers $r_1 < r_2 < \ldots < r_m$ from the sequence $1, 2, \ldots, n$ and we distinguish the terms of sum (45) for which the set $s_1, s_2, \ldots, s_m$ coincides with the set $r_1, r_2, \ldots, r_m$. This gives us part of sum (46):

$$\sum_{t_1, t_2, \ldots t_m} A\begin{pmatrix}1, & 2, & \ldots, & m\\ t_1, & t_2, & \ldots, & t_m\end{pmatrix} b_{t_1 1} b_{t_2 2} \ldots b_{t_m m}, \tag{47}$$

where summation is over all the possible permutations $(t_1, t_2, \ldots, t_m)$ of $r_1, r_2, \ldots, r_m$. On multiplying each term of (47) twice by $(-1)^{[t_1, t_2, \ldots t_m]}$, it can be shown exactly as in [6] that the sum is equal to

$$A\begin{pmatrix}1, & 2, & \ldots, & m\\ r_1, & r_2, & \ldots, & r_m\end{pmatrix} B\begin{pmatrix}r_1, & r_2, & \ldots, & r_m\\ 1, & 2, & \ldots, & m\end{pmatrix}.$$

All we need do to obtain the whole of sum (46) is to summate this product over all $r_1 < r_2 < \ldots < r_m$ which gives us (45). Finally, suppose $m > n$. In this case we can add $(m - n)$ columns of zeros to array (42) and $(m - n)$ rows of zeros to array (43). If, after this, we use the formula

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{sk} \qquad (i,\ k = 1,\ 2,\ \ldots,\ m) \tag{48}$$

instead of (44), we get the same values of c_{ik} as before, since the additional terms on the right-hand side of (48) are zero. On the other hand, arrays (42) and (43) have now become square, the corresponding determinants being zero; and it follows from the theorem for multiplying determinants that $|c_{ik}|_1^m$ is zero, so that the theorem is fully proved.

Remark. If two rectangular matrices each have m rows and n columns, multiplication of rows by rows:

$$c_{ik} = \sum_{s=1}^{n} a_{is} b_{ks} \qquad (i,\ k = 1,\ 2,\ \ldots,\ m)$$

gives us a determinant $|c_{ik}|_1^m$, the value of which is zero for $m > n$, whereas for $m \leqslant n$ it is given by

$$|c_{ik}|_1^m = \sum_{r_1<r_2<\ldots<r_m} A\begin{pmatrix}1, & 2, & \ldots, & m\\ r_1 & r_2, & \ldots, & r_m\end{pmatrix} B\begin{pmatrix}1, & 2, & \ldots, & m\\ r_1, & r_2, & \ldots, & r_m\end{pmatrix}.$$

Corollary. Let us have two square matrices of order n with elements a_{ik} and b_{ik}, whilst the numbers c_{ik} are defined by expressions (44). We express any minor $C\begin{pmatrix}p_1, p_2, \ldots, p_l\\ q_1, q_2, \ldots, q_l\end{pmatrix}$ of the determinant $|c_{ik}|_1^n$ in terms of the minors of determinants $|a_{ik}|_1^n$ and $|b_{ik}|_1^n$. It is easily seen that the square array forming the minor $C\begin{pmatrix}p_1, p_2, \ldots, p_l\\ q_1, q_2, \ldots, q_l\end{pmatrix}$ is the product of the rectangular matrices:

$$\begin{Vmatrix} a_{p_1 1}, & a_{p_1 2}, & \ldots, & a_{p_1 n}\\ a_{p_2 1}, & a_{p_2 2}, & \ldots, & a_{p_2 n}\\ \cdot & \cdot & \cdot & \cdot\\ a_{p_l 1}, & a_{p_l 2}, & \ldots, & a_{p_l n}\end{Vmatrix} \quad\text{and}\quad \begin{Vmatrix} b_{1q_1}, & b_{1q_2}, & \ldots, & b_{1q_l}\\ b_{2q_1}, & b_{2q_2}, & \ldots, & b_{2q_l}\\ \cdot & \cdot & \cdot & \cdot\\ b_{nq_1}, & b_{nq_2} & \ldots, & b_{nq_l}\end{Vmatrix}.$$

On applying the relevant theorem, we get the required expression:

$$C\begin{pmatrix}p_1, & p_2, & \ldots, & p_l\\ q_1, & q_2, & \ldots, & q_l\end{pmatrix} = \sum_{r_1<r_2<\ldots<r_l} A\begin{pmatrix}p_1, & p_2, & \ldots, & p_l\\ r_1, & r_2, & \ldots, & r_l\end{pmatrix} B\begin{pmatrix}r_1, & r_2, & \ldots, & r_l\\ q_1, & q_2, & \ldots, & q_l\end{pmatrix}, \tag{49}$$

where the r_k take their values from $1, 2, \ldots, n$. Let R_A, R_B, R_C be the ranks of matrices $\|a_{ik}\|_1^n$, $\|b_{ik}\|_1^n$, $\|c_{ik}\|_1^n$. If say $R_A < n$, and we take any $l > R_A$ in (49), all the $A\begin{pmatrix}p_1, p_2, \ldots, p_l\\ r_1, r_2, \ldots, r_l\end{pmatrix}$ will vanish by definition of R_A, so that all the $C\begin{pmatrix}p_1, p_2, \ldots, p_l\\ q_1, q_2, \ldots, q_l\end{pmatrix}$ likewise vanish. Hence it follows that $R_C < l$, i.e. $R_C \leqslant R_A$. If $\|a_{ik}\|_1^n$ is of rank n, it is obvious that $R_C \leqslant R_A$, since $R_C \leqslant n$. Similarly, $R_C \leqslant R_B$. We shall show below that if the determinant $|b_{ik}|_1^n \neq 0$, we have $R_C = R_A$, whilst if $|a_{ik}|_1^n \neq 0$, $R_C = R_B$.

§ 2. The solution of systems of equations

8. Cramer's theorem. Having described the nature and properties of determinants, we now turn to their application to the solution of systems of first degree equations. We start with the fundamental case

when the number of equations is the same as the number of unknowns. We can write such a system of n equations with n unknowns as:

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n, \end{aligned}\right\} \qquad (1)$$

the notation for the coefficients being similar to that used in [1] for the case of three equations with three unknowns. We shall make the assumption that the determinant of the system, i.e. the determinant corresponding to the array of the coefficients a_{ik}, differs from zero:

$$\Delta = |a_{ik}| \neq 0. \qquad (2)$$

We multiply both sides of the ith equation of system (1) by the cofactor of the ith element of the kth column of this determinant, i.e. both sides of the first equation are multiplied by A_{1k}, both sides of the second equation by A_{2k}, and so on. We add the equations thus obtained. The result is an equation, on the right-hand side of which we have the sum

$$A_{1k}b_1 + A_{2k}b_2 + \ldots + A_{nk}b_n,$$

whilst the coefficient of the unknown x_l on the left-hand side is given by the sum

$$A_{1k}a_{1l} + A_{2k}a_{2l} + \ldots + A_{nk}a_{nl} \quad (l = 1, 2, \ldots, n).$$

This latter sum is zero for $l \neq k$ and equal to Δ for $l = k$, i.e. we reduce to an equation of the form

$$\Delta \cdot x_k = A_{1k}b_1 + A_{2k}b_2 + \ldots + A_{nk}b_n.$$

On carrying out this procedure for each subscript k, we obtain a system of new equations as a consequence of (1):

$$\Delta \cdot x_k = A_{1k}b_1 + A_{2k}b_2 + \ldots + A_{nk}b_n \qquad (k = 1, 2, \ldots, n). \qquad (3)$$

It may easily be shown that, conversely, system (1) can be obtained as a consequence of (3). All we do is multiply both sides of the kth equation (3) by a_{lk} then sum for all k from 1 to n. We again use property V of determinants and clearly arrive at the equation

$$\Delta \cdot (a_{l1}x_1 + a_{l}x_2 + \ldots + a_{ln}x_n) = \Delta \cdot b_l \qquad (4)$$

which, after cancelling the non-zero Δ, gives us the lth equation of system (1). This procedure is possible for any l. Systems (1) and (3) are thus equivalent, and we can solve (3) instead of (1). System (3)

yields at once one and only one solution, given by

$$x_k = \frac{A_{1k}b_1 + A_{2k}b_2 + \ldots + A_{nk}b_n}{\Delta} \quad (k = 1,\ 2, \ldots, n). \tag{5}$$

We notice that, in view of our discussion in [3], the numerator of the expression written consists of the determinant obtained from Δ by replacing the elements of the kth column, i.e. the coefficients a_{ik} of x_k, by the constant terms b_i. This brings us to the following theorem.

Cramer's Theorem. *If the determinant Δ of system* (1) *differs from zero, the system has a unique solution defined by expressions* (5). *These expressions give each unknown as the quotient of two determinants, the denominator being the determinant of the system and the numerator being the determinant obtained from this by replacing the coefficients of the unknown in question by the corresponding constant terms.* Cramer's theorem is inconvenient in the case of a large number of equations with many unknowns; there are other methods that are approximate but more practical, though we shall not stop consider them.

9. The general case of systems of equations. We take the general case of m equations with n unknowns:

$$\left.\begin{aligned}
X_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k + a_{1,\,k+1}x_{k+1} + \ldots + a_{1n}x_n = b_1\\
X_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k + a_{2,\,k+1}x_{k+1} + \ldots + a_{2n}x_n = b_2\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
X_k &= a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kk}x_k + a_{k,\,k+1}x_{k+1} + \ldots + a_{kn}x_n = b_k\\
X_{k+1} &= a_{k+1,\,1}x_1 + a_{k+1,\,2}x_2 + \ldots + a_{k+1,\,k}x_k +\\
&\qquad + a_{k+1,\,k+1}x_{k+1} + \ldots + a_{k+1,\,n}x_n = b_{k+1}\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
X_m &= a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mk}x_k + a_{m,\,k+1}x_{k+1} + \ldots + a_{mn}x_n = b_m.
\end{aligned}\right\} \tag{6}$$

The complete left-hand side of the sth equation has been written X_s for the sake of brevity in later working. The coefficients a_{ik} of the system form a rectangular matrix, with rank say k. By rearranging the rows and columns, i.e. re-numbering the equations and variables, we can bring a non-zero determinant of the matrix of order k to the top left-hand corner. We call this the *leading determinant of the system.* It will have the form:

$$\Delta = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1k}\\ a_{21}, & a_{22}, & \ldots, & a_{2k}\\ \cdot & \cdot & \cdot & \cdot\\ a_{k1}, & a_{k2}, & \ldots, & a_{kk} \end{vmatrix}. \tag{7}$$

We form $(m - k)$ determinants of order $(k + 1)$ which are called *characteristic determinants of the system and which are obtained from the leading determinant by adding a row of coefficients of an equation with a number greater than k and a column of constant terms*. More precisely, the characteristic determinants are defined by the following expression:

$$\Delta_{k+s} = \begin{vmatrix} a_{11}, & a_{12}, & \dots, & a_{1k}, & b_1 \\ a_{21}, & a_{22}, & \dots, & a_{2k}, & b_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & a_{k2}, & \dots, & a_{kk}, & b_k \\ a_{k+s,1}, & a_{k+s,2}, & \dots, & a_{k+s,k}, & b_{k+s} \end{vmatrix} \tag{8}$$
$$(k + s = k + 1, \quad k + 2, \; \dots, \; m).$$

If $k = m$, i.e. the rank is equal to the number of equations, no characteristic determinants exist. We consider the further determinants obtained by replacing the last column of constant terms in a characteristic determinant by the left-hand sides of the equations:

$$\begin{vmatrix} a_{11}, & a_{12}, & \dots, & a_{1k}, & X_1 \\ a_{21}, & a_{22}, & \dots, & a_{2k}, & X_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & a_{k2}, & \dots, & a_{kk}, & X_k \\ a_{k+s,1}, & a_{k+s,2}, & \dots, & a_{k+s,k}, & X_{k+s} \end{vmatrix}. \tag{9}$$

These determinants contain x_j along with the given coefficients a_{ik}. But it is easily shown that determinants (9) are identically zero. Since

$$X_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n,$$

the last column of any one of (9) is the sum of n terms, so that, by property IV of [3], the determinant can be written as the sum of expressions of the form:

$$\begin{vmatrix} a_{11}, & a_{12}, & \dots, & a_{1k}, & a_{1j} \\ a_{21}, & a_{22}, & \dots, & a_{2k}, & a_{2j} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1} & a_{k2}, & \dots, & a_{kk}, & a_{kj} \\ a_{k+s,1}, & a_{k+s,2}, & \dots, & a_{k+s,k}, & a_{k+s,j} \end{vmatrix} . x_j .$$

The determinant appearing outside x_j is soon seen to be zero: if $j \leqslant k$, the last column is the same as one of the previous ones; whereas if $j > k$, the determinant is one of order $(k + 1)$ appearing in matrix (6)

and vanishes because the rank of (6) is k by hypothesis. On subtracting the identically zero determinants (9) from the characteristic determinants, we can write the latter as follows:

$$\Delta_{k+s} = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1k}, & b_1 - X_1 \\ a_{21}, & a_{22}, & \ldots, & a_{2k}, & b_2 - X_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & a_{k2}, & \ldots, & a_{kk}, & b_k - X_k \\ a_{k+s,1}, & a_{k+s,2}, & \ldots, & a_{k+s,k}, & b_{k+s} - X_{k+s} \end{vmatrix} \tag{10}$$
$$(k+s = k+1, \quad k+2, \ldots, m),$$

the dependence on the x_j being merely apparent in this form. We now suppose that system (6) has a solution:

$$x_1 = x_1^{(0)}, \; x_2 = x_2^{(0)}, \ldots, \; x_n = x_n^{(0)}.$$

On substituting $x_j = x_j^{(0)}$ in the last column of (10), we get a column of zeros, i.e. all the characteristic determinants must be set equal to zero.

THEOREM I. *The necessary condition for system* (6) *to have at least one solution is for all the characteristic determinants* (8) *to vanish.*

We now prove the sufficiency of the condition and give the method for finding all the solutions of the system. Thus, let all the characteristic determinants vanish. We take these in form (10) and expand by the last column. The cofactor of the element $(b_{k+s} - X_{k+s})$ is easily seen to be the leading determinant Δ which is not zero, and we can write the vanishing condition for the characteristic determinants as

$$a_1^{(k+s)}(b_1 - X_1) + a_2^{(k+s)}(b_2 - X_2) + \ldots + a_k^{(k+s)}(b_k - X_k) +$$
$$+ \Delta(b_{k+s} - X_{k+s}) = 0 \quad (k+s = k+1, \; k+2, \ldots, m), \tag{11}$$

where the $a_p^{(q)}$ are numerical coefficients of no interest to us.

Suppose now that we have a solution of the first k equations and that this is substituted for the x_j in identity (11). All the differences

$$b_1 - X_1, \quad b_2 - X_2, \ldots, \; b_k - X_k$$

now vanish, and we are left with

$$\Delta \cdot (b_{k+s} - X_{k+s}) = 0$$

or, since $\Delta \neq 0$:

$$b_{k+s} - X_{k+s} = 0 \quad (k+s = k+1, \; k+2, \ldots, m),$$

i.e. if all the characteristic determinants vanish, any solution of the first k equations must also satisfy all the remaining equations. Thus all we have to do is solve the first k equations.

We take all the unknowns with subscripts greater than k to the right-hand sides in these equations, so that they become

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k = b_1 - a_{1,\,k+1}x_{k+1} - \ldots - a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2k}x_k = b_2 - a_{2,\,k+1}x_{k+1} - \ldots - a_{2n}x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kk}x_k = b_k - a_{k,\,k+1}x_{k+1} - \ldots - a_{kn}x_n \end{array}\right\} \quad (12)$$

We consider the above as a system for determining $x_1, x_2, \ldots, x_k$. Its determinant Δ is non-zero, so that we can use Cramer's rule to obtain a determinate, unique solution. We only need to notice that the constant terms here contain $x_{k+1}, \ldots, x_n$, to which we can assign arbitrary values. Cramer's rule gives us the solution of (12) at once as:

$$x_j = \alpha_j + \beta_{k+1}^{(j)}x_{k+1} + \ldots + \beta_n^{(j)}x_n \qquad (j = 1,\ 2, \ldots, k), \qquad (13)$$

where α_s and $\beta_p^{(q)}$ are numerical coefficients and $x_{k+1}, \ \ldots, x_n$ remain arbitrary. It follows from the above that these expressions in fact give the most general solution of system (6) with the hypothesis made regarding the vanishing of all the characteristic determinants.

THEOREM II. *If all the characteristic determinants of a system vanish, only the equations containing the leading determinant need be solved, with respect to the unknowns whose coefficients make up the leading determinant. This solution can be found by Cramer's rule and expresses k unknowns, where k is the rank of the matrix of coefficients, as linear functions* (13) *of the remaining* $(n - k)$ *unknowns, the values of which remain entirely arbitrary.* All the solutions of system (6) are obtained in this way.

On comparing Theorems I and II, we arrive at the conclusion:

THEOREM III. *The necessary and sufficient condition for the existence of a solution of system* (6) *is the vanishing of all the characteristic determinants of the system.*

We remark that, if $k = n$, i.e. the rank is equal to the number of unknowns, there are no x_j whatever on the right-hand sides of (13), and all the unknowns from x_1 to x_n are fully determined.

THEOREM IV. *The necessary and sufficient condition for the system to have a unique solution is that all the characteristic determinants vanish and the rank of the matrix of coefficients is equal to the number of unknowns.*

It may be remarked that the whole of the above discussion is clearly valid for the case when the number of equations is equal to the number of unknowns, i.e. $m = n$.

Example. We take the system of four equations with three unknowns:

$$\begin{aligned} x - 3y - \ 2z &= -1 \\ 2x + \ y - \ 4z &= \ \ 3 \\ x + 4y - \ 2z &= \ \ 4 \\ 5x + 6y - 10z &= \ 10. \end{aligned}$$

We write down the matrix of coefficients:

$$\begin{Vmatrix} 1, & -3, & -2 \\ 2, & 1, & -4 \\ 1, & 4, & -2 \\ 5, & 6, & -10 \end{Vmatrix}.$$

We may easily verify that all the third order determinants in this matrix are zero, whilst the second order determinant at the top left corner differs from zero. We can thus take the latter as the leading determinant, whilst the rank of the system is two. We form the characteristic determinants, of which there are two in the present case:

$$\Delta_3 = \begin{vmatrix} 1, & -3, & -1 \\ 2, & 1, & 3 \\ 1, & 4, & 4 \end{vmatrix} = 0; \quad \Delta_4 = \begin{vmatrix} 1, & -3, & -1 \\ 2, & 1, & 3 \\ 5, & 6, & 10 \end{vmatrix} = 0.$$

Both these are zero and the given system is therefore consistent. Hence we only need to solve the first two equations with respect to x and y, z being taken to the right-hand side:

$$\begin{aligned} x - 3y &= 2z - 1 \\ 2x + y &= 4z + 3. \end{aligned}$$

The solution is obtained in the form:

$$x = \frac{\begin{vmatrix} 2z - 1, & -3 \\ 4z + 3, & 1 \end{vmatrix}}{\begin{vmatrix} 1, & -3 \\ 2, & 1 \end{vmatrix}} = 2z + \frac{8}{7}, \qquad y = \frac{\begin{vmatrix} 1, & 2z - 1 \\ 2, & 4z + 3 \end{vmatrix}}{\begin{vmatrix} 1, & -3 \\ 2, & 1 \end{vmatrix}} = \frac{5}{7},$$

z being arbitrary.

10. Homogeneous systems. A system is said to be homogeneous if all its constant terms b_i are zero. If the system has characteristic determinants, the last columns of these are made up of zeros and they

consequently vanish. Obviously, every homogeneous system has the solution

$$x_1 = x_2 = \ldots = x_n = 0$$

which we shall speak of in future as the trivial solution. The fundamental problemfor a homogeneous system is whether or not it has a non-trivial solution, and if it has, then what is the total set of such solutions? We start with the case when the number of equations is equal to the number of unknowns. The system becomes here:

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = 0 \end{array}\right\} . \tag{14}$$

If the determinant of the system differs from zero, there is a unique solution by Cramer's theorem, and this is, in fact, the trivial solution. If the determinant vanishes, the rank k of the matrix of coefficients will be less than the number n of unknowns, $(n - k)$ unknowns will thus have completely arbitrary values, and we shall have an infinite set of non-trivial solutions. Hence we arrive at the following basic theorem.

THEOREM I. *The necessary and sufficient condition for system* (14) *to have a non-trivial solution is that its determinant vanish.*

A parallel may be drawn between the results obtained for the non-homogeneous system (1) and homogeneous system (14). If the determinant of the system differs from zero, the non-homogeneous system has a unique solution whilst the homogeneous system only has the trivial solution. Whereas if the determinant vanishes, homogeneous system (14) has non-trivial solutions, yet no solution of (1) in general exists, since the existence of a solution of (1) requires a choice of constant terms such that all the characteristic determinants vanish. These parallels are of great significance below. In problems of physics, homogeneous systems are encountered when considering free vibrations and non-homogeneous systems with forced vibrations; the vanishing of the determinant for the homogeneous system characterizes the presence of proper vibrations, whereas it characterizes resonance in the case of the non-homogeneous system.

We now turn to a detailed discussion of the solutions of system (14) when its basic determinant vanishes. Let k be the rank of the matrix of its coefficients, where evidently, $k < n$. In accordance with the

theorem proved in the previous section, we have to take the k equations containing the leading determinant and solve these with respect to k unknowns. We can assume without loss of generality that these unknowns are $x_1, \ldots, x_k$. The solutions are obtained in the form:

$$x_j = \beta_{k+1}^{(j)} x_{k+1} + \ldots + \beta_n^{(j)} x_n \quad (j = 1,\ 2, \ldots,\ k), \tag{15}$$

where the $\beta_p^{(q)}$ are definite numerical coefficients and $x_{k+1}, \ldots, x_n$ can take arbitrary values.

A general property of solutions of system (14) should be noticed; this is a direct consequence of the linearity and homogeneity and may be designated the principle of superposition of solutions. If we have solutions of the system

$$x_s = x_s^{(1)};\ x_s = x_s^{(2)};\ x_s = x_s^{(3)};\ \ldots;\ x_s = x_s^{(l)}\ (s = 1,\ 2, \ldots,\ n), \tag{16}$$

further solutions are obtained by multiplying these by arbitrary constants and adding:

$$x_s = C_1 x_s^{(1)} + C_2 x_s^{(2)} + C_3 x_s^{(3)} + \ldots + C_l x_s^{(l)} \quad (s = 1,\ 2, \ldots,\ n).$$

We use the same approach as in the case of linear differential equations [II, 26] and say that solutions (16) are linearly independent if no constants C_i exist, not all zero, such that we have the equality for every s:

$$\sum_{i=1}^{l} C_i x_s^{(i)} = 0.$$

We can readily form $(n - k)$ linearly independent solutions of the system such that multiplication by arbitrary constants followed by addition gives us all the solutions. We return in fact to expressions (15) for the general solution and form solutions from these in the following manner: we put $x_{k+1} = 1$ and all the remaining x_{k+s} equal to zero in the first solution; in the second solution we put $x_{k+2} = 1$ and all the remaining x_{k+s} equal to zero, and so on; in the last, $(n - k)$th solution, we put $x_n = 1$ and all the remaining x_{k+s} equal to zero. The solutions obtained are easily seen to be linearly independent, since each contains one unknown equal to unity which is equal to zero in the remaining solutions. We denote these solutions as follows:

$$x_s = x_s^{(k+1)};\ x_s = x_s^{(k+2)};\ \ldots;\ x_s = x_s^{(m)} \qquad (s = 1,\ 2, \ldots,\ n).$$

We now take some given solution of system (14), obtained from expressions (15) with the particular values:

$$x_{k+1} = \gamma_{k+1}; \quad x_{k+2} = \gamma_{k+2}; \quad \ldots;\ x_n = \gamma_n.$$

It is clear at once that this solution is a linear combination of the solutions formed above, in fact:

$$x_s = \gamma_{k+1} x_s^{(k+1)} + \gamma_{k+2}\, x_s^{(k+2)} + \ldots + \gamma_n x_s^{(n)} \quad (s = 1, 2, \ldots, n).$$

The total number of linearly independent solutions of homogeneous system (14) is equal to $(n-k)$ for any choice of linearly independent solutions. This point will be raised again later.

We return to the general case of m homogeneous equations with n unknowns. If $m < n$, the rank k, which cannot exceed m, is likewise less than n, and $(n-k)$ unknowns remain arbitrary, i.e. *if the number of homogeneous equations is less than the number of unknowns, the system has non-trivial solutions.*

In general, $k \leqslant n$, and the system only has a trivial solution for $k = n$.

11. Linear forms. The study of systems of linear forms is closely related to the problem of solving systems of first degree equations. A linear form of the variables $x_1, x_2, \ldots, x_n$ means a linear homogeneous function of these variables. Let us have m such linear forms:

$$y_s = a_{s1}x_1 + a_{s2}x_2 + \ldots + a_{sn}x_n \qquad (s = 1, 2, \ldots, m). \tag{17}$$

These forms are said to be *linearly dependent* if there exist constants $\alpha_1, \alpha_2, \ldots, \alpha_n$, not all zero, such that we have the identity with respect to variables $x_1, x_2, \ldots, x_n$:

$$\alpha_1 y_1 + \alpha_2 y_2 + \ldots + \alpha_m y_m = 0.$$

If no such constants exist, forms (17) are said to be *linearly independent.* The coefficients of all the x_l must be equated to zero in the identity written. Hence the identity is equivalent to the following system of n equations:

$$\begin{aligned} \alpha_1 a_{11} + \alpha_2 a_{21} + \ldots + \alpha_m a_{m1} &= 0, \\ \alpha_1 a_{12} + \alpha_2 a_{22} + \ldots + \alpha_m a_{m2} &= 0, \\ \cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot & \\ \alpha_1 a_{1n} + \alpha_2 a_{2n} + \ldots + \alpha_m a_{mn} &= 0. \end{aligned}$$

The forms y_s are linearly independent when and only when this system of homogeneous equations in $\alpha_1, \alpha_2, \ldots, \alpha_m$ has only the trivial solution.

The results obtained above lead to a number of conclusions regarding the linear dependence of forms. If $m > n$, the homogeneous system written certainly has non-trivial solutions and the forms are

linearly dependent. The necessary and sufficient condition for the forms to be independent is that the rank k of the matrix of coefficients a_{pq} is equal to the number of forms m. *If $m = n$, i.e. the number of forms is equal to the number of variables, the necessary and sufficient condition for linear independence is the non-vanishing of the total square ($m = n$) matrix of a_{pq}.* We speak in this case of the existence of a *complete system of linearly independent forms.* If $m \leqslant n$ and forms (17) are linearly independent (i.e. $k = m$), the system of equations (17) is soluble for any values of y_s with respect to the variables x_t whose coefficients form a non-zero determinant of order k, i.e. *linearly independent forms can take any set of values y_s.* If $k = m = n$, all the variables x_t are defined for given y_s.

We now take $k < m$. By suitable numbering of the forms y_s and variables x_t, we can arrange for a non-zero determinant of order k to stand at the top left corner of the matrix of a_{pq}. With this, the first k forms, $y_1, y_2, \ldots, y_k$ are linearly independent, whilst each of the remaining form y_{k+t} can be expressed linearly in terms of the first k forms. This follows because the rank, equal to k, of the matrix of coefficients of the first k forms is the same as the number of forms, whence their linear independence. If we take $(k + 1)$ forms $y_1, y_2, \ldots,$ y_k, y_{k+t}, the rank of the matrix of their coefficients is still k and is less than the number of forms, i.e. the forms are linearly dependent, so that there exist constants β_s such that

$$\beta_1 y_1 + \ldots + \beta_k y_k + \beta_{k+1} y_{k+1} = 0.$$

The coefficient β_{k+1} in this relationship must differ from zero, since otherwise the first k forms would be linearly dependent. Hence we have a linear expression for y_{k+1} in terms of the first k forms:

$$y_{k+1} = -\frac{\beta_1}{\beta_{k+1}} y_1 - \frac{\beta_2}{\beta_{k+1}} y_2 - \ldots - \frac{\beta_k}{\beta_{k+1}} y_k .$$

The number k is called *the rank of the system of forms* (17). *This number is equal to the rank of the matrix of coefficients on the one hand, and on the other, to the greatest number of linearly independent forms of system* (17).

Suppose we have k linearly independent forms $y_1, y_2, \ldots, y_k$, where $k < n$. We can assume that the kth order determinant at the top left corner of the matrix of a_{pq} differs from zero. This system of k forms may easily be extended to become a complete system of n linearly independent forms. All we need do is take say

$$y_{k+1} = x_{k+1}; \quad \ldots ; \quad y_n = x_n .$$

The determinant of the n forms obtained will be:

$$\begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1k}, & a_{1,k+1}, & & \ldots, & a_{1n} \\ a_{21}, & a_{22}, & \ldots, & a_{2k}, & a_{2,k+1}, & & \ldots, & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & a_{k2}, & \ldots, & a_{kk}, & a_{k,k+1}, & & \ldots, & a_{kn} \\ 0, & 0 & \ldots, & 0, & 1, & 0, & \ldots, & 0 \\ 0, & 0 & \ldots, & 0, & 0, & 1, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0 & \ldots, & 0, & 0, & 0, & \ldots, & 1 \end{vmatrix}.$$

On expanding this determinant by the last row, then the next to the last row, and so on, we see that its magnitude is equal to the kth order determinant at the top left corner, i.e. is non-zero. Thus the forms $y_1, y_2, \ldots, y_n$ are in fact linearly independent. It follows that *every system of linearly independent forms can be extended to become a complete system of linearly independent forms.*

12. n-dimensional vector space. The results obtained above are open to a geometrical interpretation which will be useful later. We introduce for this purpose the concept of a vector in n-dimensional space, a vector being defined as a set of n (complex) numbers appearing in a definite order. *Any such vector* $\mathbf{x}$ *is characterized by a sequence of n complex numbers, known as the components of the vector:* $\mathbf{x}(x_1, x_2, \ldots, x_n)$. *The aggregate of all these vectors forms an n-dimensional vector space* R_n.

Two vectors are taken to be equal when and only when all their components are the same, i.e. if $\mathbf{u}(u_1, u_2, \ldots, u_n)$ and $\mathbf{v}(v_1, v_2, \ldots, v_n)$ are two vectors, the vector equation $\mathbf{u} = \mathbf{v}$ is equivalent to the following scalar equations: $u_1 = v_1$; $u_2 = v_2$; $\ldots$, $u_n = v_n$. We next define multiplication of a vector by a number and addition of vectors. Multiplication of a vector by a number amounts by definition to multiplication of all the components of the vector by the number, i.e. if vector $\mathbf{x}$ has components $(x_1, x_2, \ldots, x_n)$, vector kx has components $(kx_1, kx_2, \ldots, kx_n)$. Addition of vectors amounts to addition of their components, i.e. if we have vectors $\mathbf{x}(x_1, x_2, \ldots, x_n)$ and $\mathbf{y}(y_1, y_2, \ldots, y_n)$, their sum $\mathbf{x} + \mathbf{y}$ has by definition components $(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$. The null vector is defined as the vector $(0, 0, \ldots, 0)$, all the components of which are zero. We write

the null vector as θ. We obviously have $\theta = 0\mathbf{x}$, where $\mathbf{x}$ is any vector, and $\mathbf{x} + \theta = \mathbf{x}$. Subtraction of vectors is defined thus: the vector $\mathbf{x} - \mathbf{y}$ has components $(x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n)$. Obviously, $\mathbf{x} - \mathbf{x} = \theta$ and $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-1)\mathbf{y}$, i.e. subtraction of vector $\mathbf{y}$ is equivalent to addition of $\mathbf{y}$ multiplied by (-1). We shall often have to write vector equations below. Any such equation is equivalent to n scalar equations, expressing the fact that corresponding components of each side are equal. Though we shall not use the symbol θ below for the null vector, it must be borne in mind that a zero appearing on one side of an equation is to be read as the null vector. The ordinary properties of addition and multiplication follow at once from the definitions given above:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}; \quad \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z};$$

$$(k_1 + k_2)\,\mathbf{x} = k_1\mathbf{x} + k_2\mathbf{x}; \quad k\,(\mathbf{x} + \mathbf{y}) = k\mathbf{x} + k\mathbf{y}; \quad k_1\,(k_2\mathbf{x}) = (k_1k_2)\,\mathbf{x}.$$

We can thus transpose or group together the terms in a vector sum with any number of terms. From the equation $\mathbf{x} + \mathbf{y} = \mathbf{z}$, it follows that $\mathbf{x} = \mathbf{z} - \mathbf{y}$ and $\mathbf{y} = \mathbf{z} - \mathbf{x}$, and conversely, from $\mathbf{x} - \mathbf{y} = \mathbf{z}$ it follows that $\mathbf{x} = \mathbf{y} + \mathbf{z}$.

We now introduce the concepts of linear dependence and independence for vectors. *The vectors*

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(l)} \tag{18}$$

will be said to be linearly dependent if there exist constants $C_1, \ldots, C_l$, not all zero, such that

$$C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)} + \ldots + C_l\mathbf{x}^{(l)} = 0. \tag{19}$$

If no such constants exist, vectors (18) *are said to be linearly independent.* We write the components of vectors $\mathbf{x}^{(j)}$ as $(x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)})$. Condition (19) is clearly equivalent to the system of n equations with unknowns $C_1, C_2, \ldots, C_l$:

$$\left.\begin{aligned} x_1^{(1)}C_1 + x_1^{(2)}C_2 + \ldots + x_1^{(l)}C_l &= 0 \\ x_2^{(1)}C_1 + x_2^{(2)}C_2 + \ldots + x_2^{(l)}C_l &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n^{(1)}C_1 + x_n^{(2)}C_2 + \ldots + x_n^{(l)}C_l &= 0. \end{aligned}\right\} \tag{20}$$

By using the results obtained above for homogeneous systems, we can easily draw a number of conclusions and interpret them geometrically. Let us first take $l > n$, i.e. the number of vectors is greater than the number of spatial dimensions. With this, the number

of equations in the homogeneous system (20) is less than the number of unknowns, and, as we know, the system certainly has non-zero solutions for the C_j, i.e. the vectors are certainly linearly dependent. In other words, *the number of linearly independent vectors is at most equal to the number of dimensions*. We now take the case $l = n$. Here system (20) contains as many equations as unknowns and has non-zero solutions when and only when its determinant vanishes, i.e. if we have n vectors in n-dimensional space, and form a determinant from the n^2 components, locating say the components of a given vector in a given column, with the rows having the same numbering as the components, the necessary and sufficient condition for linear independence of the vectors is the non-vanishing of this determinant. The magnitude of the determinant is analogous to the volume of a parallelepiped in real three-dimensional space.

We can consider the elements $(b_{1k}, b_{2k}, \ldots, b_{nk})$ of each column in any determinant $| b_{ik} |$ of order n as the components of a vector $\mathbf{b}^{(k)}$, the magnitude of the determinant being here a function of the n vectors $\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(n)}$. The vanishing of the determinant is equivalent to the fact that the vectors are linearly dependent.

The magnitude of the determinant, considered as a function of vectors $\mathbf{b}^{(k)}$, is written

$$|b_{ik}| = \Delta\,(\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \ldots, \mathbf{b}^{(n)}).$$

On recalling that the magnitude of a determinant changes sign on interchange of two columns, we can say that the function Δ merely changes sign on interchange of two of its arguments. Such a function is usually said to be *anti-symmetric*. It may readily be seen that for instance the Vandermonde determinant D_n, considered in [5], is likewise an anti-symmetric function of its arguments $x_1, \ldots, x_n$.

We return to system (20) and the question of the linear independence of vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(l)}$, on the assumption that $l \leqslant n$. Let k be the rank of the matrix formed by the components $x_p^{(q)}$. If $k = l$, as we saw, the system has only the trivial solution, i.e. the vectors are linearly independent. Whereas if $k < l$, the system certainly has a non-trivial solution, i.e. *the necessary and sufficient condition for vectors to be linearly independent is for their number to be equal to the rank of the matrix formed by their components*. We now assume $k < l$, i.e. the vectors are linearly dependent. We distinguish among these the k vectors whose components contain a kth order non-zero determinant

(there may be more than one way of doing this). By what was proved above, these k vectors are linearly independent. Each of the remaining vectors is easily seen to be linearly expressible in terms of the chosen vectors. In fact, let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ be the linearly independent vectors. On associating with these any vector $\mathbf{x}^{(k+s)}$, we get $(k+1)$ vectors, which will be linearly dependent, since the rank k of the matrix of their components is less than the number $l = k + 1$. Hence constants C_i $(i = 1, 2, \ldots, k, k + s)$ will exist, not all zero, such that

$$C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)} + \ldots + C_k\mathbf{x}^{(k)} + C_{k+s}\mathbf{x}^{(k+s)} = 0.$$

We certainly have $C_{k+s} \neq 0$ here, since otherwise vectors $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$ would be linearly dependent. Hence the equation gives us

$$\mathbf{x}^{(k+s)} = -\frac{C_1}{C_{k+s}}\mathbf{x}^{(1)} - \frac{C_2}{C_{k+s}}\mathbf{x}^{(2)} - \ldots - \frac{C_k}{C_{k+s}}\mathbf{x}^{(k)},$$

i.e. $\mathbf{x}^{(k+s)}$ is expressed linearly in terms of $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$. Let $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$ be any n linearly independent vectors. We can take as an example of these:

$$(1, 0, 0, \ldots 0); \quad (0, 1, 0, \ldots 0); \quad \ldots; \quad (0, 0, 0, \ldots, 1). \qquad (21)$$

If we take any vector we please, $\mathbf{x}$, the $(n + 1)$ vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$, $\mathbf{x}^{(n)}$, $\mathbf{x}$ are in fact linearly dependent, as we have seen:

$$C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)} + \ldots + C_n\mathbf{x}^{(n)} + C\mathbf{x} = 0,$$

the constant C being unquestionably non-zero, since otherwise vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ would be linearly dependent. It follows from the above that *any vector* $\mathbf{x}$ *is expressible linearly in terms of* n *linearly independent vectors:*

$$\mathbf{x} = a_1\mathbf{x}^{(1)} + a_2\mathbf{x}^{(2)} + \ldots + a_n\mathbf{x}^{(n)} \qquad \left(a_s = -\frac{C_s}{C}\right). \qquad (22)$$

It may easily be seen that *the expression for* x *in terms of* $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$, $\mathbf{x}^{(n)}$ *is unique.* If, in addition to the above expression, there existed the further expression

$$\mathbf{x} = \beta_1\mathbf{x}^{(1)} + \beta_2\mathbf{x}^{(2)} + \ldots + \beta_n\mathbf{x}^{(n)},$$

where the β_s differ from the corresponding a_s, subtraction of the two expressions would give us

$$(a_1 - \beta_1)\mathbf{x}^{(1)} + (a_2 - \beta_2)\mathbf{x}^{(2)} + \ldots + (a_n - \beta_n)\mathbf{x}^{(n)} = 0,$$

i.e. vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ are linearly dependent, which is false. If we take vectors (21) for $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$, the a_s in (22) are evidently the

components x_s of the vector $\mathbf{x}(x_1, x_2, \ldots, x_n)$. We can also speak of the a_s as *the components of* $\mathbf{x}$ *in the general case, when* $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}$ *are taken as the fundamental vectors.* On assigning all possible complex values to the numbers a_s, we obtain all the vectors of our n-dimensional space. We now suppose that we have k linearly independent vectors

$$\mathbf{x}^{(1)},\ \mathbf{x}^{(2)},\ \ldots,\ \mathbf{x}^{(k)}, \tag{23}$$

where $k < n$. The set of vectors obtained in accordance with

$$\mathbf{y} = C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)} + \ldots + C_k\mathbf{x}^{(k)}, \tag{23_1}$$

where the C_s are arbitrary constants, is said to form a *k-dimensional subspace L_k*. It can be shown as above that any vector belonging to L_k is uniquely expressible in terms of $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$. In other words, *vectors* (23) *form a subspace L_k*.

We notice that, if any vector $\mathbf{z}$ belongs to L_k, i.e. is expressible by an equation of type (23_1), the vector $c\mathbf{z}$, where c is any constant, is evidently also given by an equation of type (23_1), i.e. also belongs to L_k. Similarly, if $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ belong to L_k, their sum $\mathbf{z}^{(1)} + \mathbf{z}^{(2)}$ also belongs to L_k. Hence a more general property follows at once: *if vectors $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(p)}$ belong to L_k, any linear combination of them, $\gamma_1 \mathbf{z}^{(1)} + \gamma_2 \mathbf{z}^{(2)} + \ldots + \gamma_p \mathbf{z}^{(p)}$ also belongs to L_k.*

We take any m vectors belonging to L_k:

$$\mathbf{y}^{(s)} = C_1^{(s)}\mathbf{x}^{(1)} + C_2^{(s)}\mathbf{x}^{(2)} + \ldots + C_k^{(s)}\mathbf{x}^{(k)} \qquad (s = 1,\ 2,\ \ldots,\ m). \tag{24}$$

In view of the linear independence of vectors (23), a relationship of the form

$$a_1\mathbf{y}^{(1)} + a_2\mathbf{y}^{(2)} + \ldots + a_m\mathbf{y}^{(m)} = 0$$

is equivalent to a system of k homogeneous equations in $a_1, a_2, \ldots, a_k$:

$$a_1C_q^{(1)} + a_2C_q^{(2)} + \ldots + a_mC_q^{(m)} = 0 \qquad (q = 1,\ 2,\ \ldots,\ k).$$

If this system has a non-trivial solution, vectors (24) are linearly dependent. In particular, if $m > k$, non-trivial solutions certainly exist, i.e. any set of more than k vectors of the subspace formed by vectors (23) is a linearly dependent set. It follows at once from this that the subspace formed by linearly independent vectors (23) cannot be formed by using a set of linearly independent vectors $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(l)}$, the number of which is $l < k$. For otherwise, by what we have proved above, there could not exist more than l linearly independent vectors in the subspace, whilst on the other hand, the linearly independent vectors (23), the number of which, k, is greater than l, have to belong

to the subspace. *If we take any k linearly independent vectors $\mathbf{u}^{(1)}$, $\mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(k)}$, belonging to L_k, they form, in the sense indicated above, the same subspace L_k.*

This follows because, by definition of subspace, any linear combination

$$C_1\mathbf{u}^{(1)} + C_2\mathbf{u}^{(2)} + \ldots + C_k\mathbf{u}^{(k)}$$

belongs to L_k. Whereas if we take any vector $\mathbf{y}$ of L_k, the $(k+1)$ vectors $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}, \mathbf{y}$ belong to L_k and are therefore linearly dependent, by the above:

$$\beta_1\mathbf{u}^{(1)} + \beta_2\mathbf{u}^{(2)} + \ldots + \beta_k\mathbf{u}^{(k)} + \gamma\mathbf{y} = 0,$$

where γ must be non-zero since $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$ are linearly independent. This gives us the result that any vector $\mathbf{y}$ of L_k can be expressed in terms of $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(k)}$, i.e. these latter vectors in fact produce L_k. If $m = k$ in expressions (24), and the determinant of the coefficients $C_p^{(q)}$ differs from zero, $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ must be linearly independent vectors of L_k. It is easily shown that in general the number of linearly independent vectors yielded by expressions (24) is equal to the rank of the matrix of $C_p^{(q)}$.

We saw above that if a vector $\mathbf{z}$ belongs to a certain subspace L, the vector $c\mathbf{z}$, where c is any constant, also belongs to L; and if $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ belong to L, their sum also belongs to L. We might have given another definition of subspace, viz, *a subspace is a set of vectors such that, if $\mathbf{z}$ belongs to L, $c\mathbf{z}$ belongs to L, whilst if $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$ belong to L, $(\mathbf{z}^{(1)} + \mathbf{z}^{(2)})$ also belongs to L.* An immediate consequence of this is that any linear combination of vectors belonging to L also belongs to L. We have just seen that the properties forming part of the new definition follow as corollaries from the first definition. We can show conversely that the first is a consequence of the new definition, i.e. *the two definitions are equivalent.*

Let $\mathbf{x}^{(1)}$ be a certain vector belonging to L. By definition of L, vectors $C_1\mathbf{x}^{(1)}$, with arbitrary C_1, also belong to L. If L is altogether exhausted by these vectors, we must have L_1 in the previous sense. If this is not the case, and a vector $\mathbf{x}^{(2)}$, linearly independent of $\mathbf{x}^{(1)}$, appears in L, the vectors $C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)}$, with arbitrary C_1 and C_2, belong to L. If L is altogether exhausted by these vectors, L is an L_2 in the previous sense. If the opposite is the case, an $\mathbf{x}^{(3)}$ appears in L, such that $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}$ are linearly independent. By proceeding in this way, we can exhaust L completely by means of a finite set of linearly independent vectors, the number of these being not greater than n.

The greatest number k of these linearly independent $\mathbf{x}^{(s)}$ gives us the dimensions of the subspace L. If it happens that $k = n$, L coincides with the total n-dimensional space.

We note a point in connection with the formation of subspaces. Let the vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$ be linearly dependent. We can now say, as before, that formula (23_1) defines a subspace L. Let the first l vectors: $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(l)}$, be linearly independent, whilst each of the remaining vectors: $\mathbf{x}^{(l+1)}, \ldots, \mathbf{x}^{(k)}$ can be expressed linearly in terms of the first l. The set of vectors defined by (23_1) is now clearly the same as the set defined by

$$\mathbf{y} = C_1\mathbf{x}^{(1)} + C_2\mathbf{x}^{(2)} + \ldots + C_l\mathbf{x}^{(l)},$$

i.e. the subspace L defined by the linearly dependent $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$, is l-dimensional ($l < k$).

Let us take real three-dimensional space and agree to measure vectors from a fixed point O (the origin). Here, $n = 3$. With $k = 1$, the subspace L_1 is a straight line passing through O, whilst L_2 is a plane passing through O.

13. Scalar product. We use the following notation: if a is a complex number, $\bar{a}$ is the complex conjugate of a, and $|a|$ is the modulus of a. We thus have $a\bar{a} = |a|^2$. If a is real, $\bar{a} = a$ and $|a|^2 = a^2$. We now introduce a new concept, of great importance for what follows.

DEFINITION. *The scalar product of two vectors*

$$\mathbf{x}\,(x_1\ x_2, \ldots, x_n) \quad \text{and} \quad \mathbf{y}\,(y_1\ y_2, \ldots, y_n)$$

is defined as the number represented by the sum

$$\sum_{s=1}^{n} x_s\bar{y}_s.$$

We shall denote the scalar product by the symbol $(\mathbf{x}, \mathbf{y})$. We have:

$$(\mathbf{x}, \mathbf{y}) = \sum_{s=1}^{n} x_s\,\bar{y}_s; \qquad (\mathbf{y}, \mathbf{x}) = \sum_{s=1}^{n} y_s\,\bar{x}_s,$$

whence it follows that

$$(\mathbf{y}, \mathbf{x}) = \overline{(\mathbf{x}, \mathbf{y})}\,.$$

We say that *two vectors are perpendicular or orthogonal to each other if their scalar product is zero*. Inasmuch as the conjugate of zero is zero, the order of the vectors in the scalar product has no importance as

regards the condition for orthogonality. Obviously, the null vector $(0, 0, \ldots, 0)$ is orthogonal to any vector $\mathbf{x}$.

The properties

$$(a\mathbf{x}, \mathbf{y}) = a(\mathbf{x}, \mathbf{y}); \quad (\mathbf{x}, a\mathbf{y}) = \bar{a}(\mathbf{x}, \mathbf{y}),$$

where a is a numerical factor, follow at once from the definition of scalar product. Furthermore:

$$(\mathbf{x} + \mathbf{y}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) + (\mathbf{y}, \mathbf{z}); \quad (\mathbf{x}, \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{z}),$$

the distributive law being valid for any number of terms. We have, for instance:

$$(\mathbf{x} + \mathbf{y}, \mathbf{u} + \mathbf{v}) = (\mathbf{x}, \mathbf{u}) + (\mathbf{x}, \mathbf{v}) + (\mathbf{y}, \mathbf{u}) + (\mathbf{y}, \mathbf{v}).$$

We form the scalar product of $\mathbf{x}(x_1, x_2, \ldots, x_n)$ with itself:

$$(\mathbf{x}, \mathbf{x}) = \sum_{s=1}^{n} x_s \bar{x}_s = \sum_{s=1}^{n} |x_s|^2.$$

We thus get a real number, positive for non-zero vectors $\mathbf{x}$, and zero for the null vector $(0, 0, \ldots, 0)$. *The square root (numerical value) of the real number* $(\mathbf{x}, \mathbf{x})$ *is called the norm or length of vector* $\mathbf{x}$. On using $\|\mathbf{x}\|$ to denote the norm, we can write:

$$\|\mathbf{x}\|^2 = (\mathbf{x}, \mathbf{x}) = \sum_{s=1}^{n} |x_s|^2; \quad \|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{\sum_{s=1}^{n} |\mathbf{x}_s|^2}.$$

The equation $\|\mathbf{x}\| = 0$ is equivalent to the fact that $\mathbf{x}$ is the null vector. Suppose we have three mutually perpendicular vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, i.e.

$$(\mathbf{x}, \mathbf{y}) = 0; \quad (\mathbf{x}, \mathbf{z}) = 0; \quad (\mathbf{y}, \mathbf{z}) = 0.$$

On using the distributive law for scalar products and taking into account the equations written, we get:

$$(\mathbf{x} + \mathbf{y} + \mathbf{z}, \mathbf{x} + \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z})$$

or

$$\|\mathbf{x} + \mathbf{y} + \mathbf{z}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2.$$

This expresses *the theorem of Pythagoras*. It is valid for any number of terms, with the essential proviso that the terms are orthogonal in pairs. We show that *if the vectors* $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(l)}$, *none of which is the null vector, are orthogonal in pairs, they are linearly independent.* We take

$$\sum_{s=1}^{l} C_s \mathbf{x}^{(s)} = 0$$

and show that all the numbers C_s must be zero. We form the scalar product of both sides of this equation with $\mathbf{x}^{(k)}$, where k is one of the numbers $1, 2, \ldots, l$:

$$\sum_{s=1}^{l} C_s (\mathbf{x}^{(s)}, \mathbf{x}^{(k)}) = 0 .$$

Since pairs of the $\mathbf{x}^{(s)}$ are orthogonal, we have $(\mathbf{x}^{(s)}, \mathbf{x}^{(k)}) = 0$ for $s \neq k$; hence the above equation gives: $C_k(\mathbf{x}^{(k)}, \mathbf{x}^{(k)}) = 0$, i.e. $C_k \| \mathbf{x}^{(k)} \|^2 = 0$, whence, since $\| \mathbf{x}^{(k)} \|^2 > 0$, it follows that $C_k = 0$, this being true for any choice of k.

14. Geometrical interpretation of homogeneous systems. We take the homogeneous system

$$\left.\begin{array}{l} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n = 0 \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1} x_1 + a_{n2} x_2 + \ldots + a_{nn} x_n = 0. \end{array}\right\} \tag{25}$$

We bring in the vectors

$$\mathbf{a}^{(1)} (\bar{a}_{11}, \bar{a}_{12}, \ldots, \bar{a}_{1n}); \ \ldots; \ \mathbf{a}^{(n)} (\bar{a}_{n1}, \bar{a}_{n2}, \ldots, \bar{a}_{nn}). \tag{26}$$

System (25) can now be written in the compressed form:

$$(\mathbf{x}, \mathbf{a}^{(1)}) = 0; \ \ldots; \ (\mathbf{x}, \mathbf{a}^{(n)}) = 0, \tag{27}$$

so that the problem amounts to finding a vector $\mathbf{x}$, perpendicular to all the vectors $\mathbf{a}^{(j)}$. If the determinant $| a_{ik} |$ differs from zero, the determinant $| \bar{a}_{ik} |$, with conjugate magnitude, also differs from zero. In this case, the vectors $\mathbf{a}^{(j)}$ are linearly independent, and system (27) only has a trivial solution, i.e. there exists no vector (apart from the null vector) which is simultaneously perpendicular to n linearly independent vectors (in n-dimensional space).

We now take the case when the determinant of system (25) vanishes. Let the rank of the system be k. If a matrix is formed of the conjugate elements, the determinants appearing in it will be conjugate in magnitude to the determinants appearing in the array of a_{ik}, and the rank of the conjugate matrix will also evidently be k. Hence, by what we have shown above, there will be k linearly independent vectors among the $\mathbf{a}^{(j)}$, the remainder being linear combinations of these. We can suppose without loss of generality that these linearly independent vectors are

$$\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(k)}, \tag{28}$$

whilst for the remainder we have expressions of the form

$$\mathbf{a}^{(k+s)} = \beta_1^{(k+s)}\mathbf{a}^{(1)} + \ldots + \beta_k^{(k+s)}\mathbf{a}^{(k)} \quad (k+s = k+1,\ k+2, \ldots, n),$$

where the $\beta_p^{(q)}$ are numerical coefficients. It now follows at once that if $\mathbf{x}$ is orthogonal to vectors (28), it is perpendicular to all the vectors $\mathbf{a}^{(j)}$. In fact:

$$(\mathbf{x}, \mathbf{a}^{(k+s)}) = \sum_{i=1}^{k} (\mathbf{x}, \bar{\beta}_i^{(k+s)}\mathbf{a}^{(i)})$$

and the sum as a whole is zero since each individual term vanishes by hypothesis. It is thus sufficient to solve the first k equations of the system. Assuming, as usual, that a non-zero determinant of order k is at the top left corner, we get $(n-k)$ linearly independent solutions $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n-k)}$ for the required vector $\mathbf{x}$ by the method indicated in [12], and every solution will consist of a linear combination of these $(n-k)$ vectors. We can say in our present case that the vectors given by

$$\mathbf{y} = C_1\mathbf{a}^{(1)} + \ldots + C_k\mathbf{a}^{(k)},$$

where the C_i are arbitrary constants, form a k-dimensional space L_k which is in fact a subspace of the total n-dimensional space. In the same way, the vectors obtained, $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n-k)}$, form an $(n-k)$-dimensional subspace M_{n-k}. *The subspace M_{n-k} is orthogonal to the subspace L_k* in the sense that any vector of M_{n-k} is orthogonal to any vector of L_k (and conversely, of course). The subspace M_{n-k} consists of the vectors which satisfy system (27), i.e. are orthogonal to $\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(k)}$. The n vectors $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(k)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n-k)}$ are readily seen to be linearly independent. For suppose, on the contrary, that a relationship exists between them:

$$(c_1\mathbf{a}^{(1)} + \ldots + c_k\mathbf{a}^{(k)}) + (d_1\mathbf{x}^{(1)} + \ldots + d_{n-k}\mathbf{x}^{(n-k)}) = 0. \quad (29)$$

The first bracket yields a vector $\mathbf{a}$ of L_k, and the second a vector $\mathbf{x}$ of M_{n-k}, and we now have $\mathbf{a} + \mathbf{x} = 0$ or $\mathbf{a} = -\mathbf{x}$. But $\mathbf{a}$ and $\mathbf{x}$ are orthogonal to each other, i.e. $\mathbf{a}$ must be orthogonal to itself, in other words, $(\mathbf{a}, \mathbf{a}) = 0$ or $\mathbf{a} = 0$ which means that $\mathbf{a}$ is the null vector. The same can be said of $\mathbf{x}$. Hence:

$$c_1\mathbf{a}^{(1)} + \ldots + c_k\mathbf{a}^{(k)} = 0 \quad \text{and} \quad d_1\mathbf{x}^{(1)} + \ldots + d_{n-k}\mathbf{x}^{(n-k)} = 0.$$

But $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(k)}$ are linearly independent by hypothesis, and all the constants c_s must consequently vanish; and the same can be said as regards the d_s. All the coefficients in (29) therefore vanish, i.e.

vectors $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(k)}\ \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n-k)}$ are in fact linearly independent.

Every vector $\mathbf{x}$ can be uniquely represented in the form

$$\mathbf{x} = (\gamma_1 \mathbf{a}^{(1)} + \ldots + \gamma_k \mathbf{a}^{(k)}) + (\delta_1 \mathbf{x}^{(1)} + \ldots + \delta_{n-k} \mathbf{x}^{(n-k)}),$$

the first bracket giving the vector belonging to L_k and the second the vector belonging to M_{n-k}. As we have already mentioned, the vectors that compose M_{n-k} are all the possible solutions of system (27), and hence, whatever the choice of the total system of linearly independent solutions, the number of these solutions is equal to $(n - k)$, i.e. equal to the number of dimensions of M_{n-k}. The earlier discussion of homogeneous systems leads to the following important result.

If L_k is a k-dimensional subspace ($k < n$), the vectors orthogonal to it form an $(n - k)$-dimensional subspace M_{n-k}, and every vector $\mathbf{x}$ of R_n can be written as the sum $\mathbf{x} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y}$ belongs to L_k and $\mathbf{z}$ to M_{n-k}.

We show that the representation of $\mathbf{x}$ as a sum is unique. Suppose that, in addition to the above, we have $\mathbf{x} = \mathbf{u} + \mathbf{v}$, where $\mathbf{u}$ belongs to L_k and $\mathbf{v}$ to M_{n-k}. We want to show that $\mathbf{u} = \mathbf{y}$ and $\mathbf{v} = \mathbf{z}$. We have: $\mathbf{y} + \mathbf{z} = \mathbf{u} + \mathbf{v}$, whence $\mathbf{y} - \mathbf{u} = \mathbf{v} - \mathbf{z}$. The difference $\mathbf{y} - \mathbf{u}$ belongs to L_k, whilst $\mathbf{v} - \mathbf{z}$ belongs to M_{n-k}, whence it follows that $\mathbf{y} - \mathbf{u}$ is orthogonal to itself, i.e. $(\mathbf{y} - \mathbf{u}, \mathbf{y} - \mathbf{u}) = 0$ or $\| \mathbf{y} - \mathbf{u} \| = 0$, so that $\mathbf{y} - \mathbf{u} = 0$ and $\mathbf{y} = \mathbf{u}$. Since $\mathbf{y} - \mathbf{u} = \mathbf{v} - \mathbf{z}$, it now follows that $\mathbf{v} = \mathbf{z}$. In the representation of $\mathbf{x}$ as $\mathbf{x} = \mathbf{y} + \mathbf{z}$, $\mathbf{y}$ is known as *the projection of* $\mathbf{x}$ *on the subspace* L_k. The vectors $\mathbf{y}$ and $\mathbf{z}$ are orthogonal, and Pythagoras' theorem gives: $\|\mathbf{x}\|^2 = \|\mathbf{y}\|^2 + \|\mathbf{z}\|^2$, whence we have $\|\mathbf{y}\| \leqslant \|\mathbf{x}\|$, the sign of equality being obtained when and only when $\mathbf{z}$ is the null vector, i.e. when $\mathbf{x}$ belongs to L_k, so that $\mathbf{y} = \mathbf{x}$. Similarly, $\|\mathbf{z}\| \leqslant \|\mathbf{x}\|$, and the sign of equality is obtained when and only when $\mathbf{x}$ is orthogonal to L_k, i.e. $\mathbf{z} = \mathbf{x}$. We usually describe L_k and M_{n-k} as *complementary orthogonal subspaces*. If $k = n$, L_n is the whole of R_n, whilst M_0 reduces to the null vector.

Let us take real three-dimensional space that we discussed above, and let $k = 2$, so that $n - k = 3 - 2 = 1$. The subspace L_2 is a plane P, passing through the point O, whilst M_1 is a straight line passing through O and perpendicular to P. Any vector can be uniquely represented as the sum of two vectors, one of which lies in the plane P, whilst the other is along the line K. We have interpreted geometrically the solution of a homogeneous system in the case when the number of equations is equal to the number of unknowns. The general case can be treated in precisely the same way, when the number of vectors $\mathbf{a}^{(s)}$

is not necessarily equal to n. Similar remarks apply as regards the next article.

15. Non-homogeneous systems. We take the non-homogeneous system:

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + n_{n2}x_2 + \dots + a_{nn}x_n = b_n. \end{array}\right\} \quad (30)$$

This can be interpreted as the problem of finding the vector $\mathbf{x}(x_1, x_2, \dots, x_n)$ from the system:

$$(\mathbf{x}, \mathbf{a}^{(1)}) = b_1; \dots; (\mathbf{x}, \mathbf{a}^{(n)}) = b_n. \quad (31)$$

given the vectors (26).

If the determinant of the system differs from zero, Cramer's theorem provides a unique solution. Suppose the determinant vanishes and the rank of the matrix of its coefficients is k, a non-zero determinant of order k being situated at the top left corner as usual. Along with system (30), we write down the system of homogeneous equations whose coefficients are obtained from the coefficients of the given system by replacing rows with columns and all numbers with their conjugates. The system will take the form:

$$\left.\begin{array}{l} \bar{a}_{11}y_1 + \bar{a}_{21}y_2 + \dots + \bar{a}_{n1}y_n = 0 \\ \bar{a}_{12}y_1 + \bar{a}_{22}y_2 + \dots + \bar{a}_{n2}y_n = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \bar{a}_{1n}y_1 + \bar{a}_{2n}y_2 + \dots + \bar{a}_{nn}y_n = 0. \end{array}\right\} \quad (32)$$

As before, the matrix of its coefficients has rank k and a non-zero kth order determinant stands at the top left corner. The homogeneous system is known as the *adjoint* of system (30). We have seen above that its general solution is a linear combination of $(n - k)$ solutions (vectors) which can be obtained, for instance, by using Cramer's theorem to solve the first k equations with respect to $y_1, \dots, y_k$, the remaining y_{k+s} being put equal to zero except for one which is put equal to unity. This method brings us, with $y_{k+1} = 1$, to the system:

$$\begin{array}{l} \bar{a}_{11}y_1 + \bar{a}_{21}y_2 + \dots + \bar{a}_{k1}y_k = -\bar{a}_{k+1,1}, \\ \bar{a}_{12}y_1 + \bar{a}_{22}y_2 + \dots + \bar{a}_{k2}y_k = -\bar{a}_{k+1,2}, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \bar{a}_{1k}y_1 + \bar{a}_{2k}y_2 + \dots + \bar{a}_{kk}y_k = -\bar{a}_{k+1,k}. \end{array}$$

On solving this system and taking conjugate values, we get:

$$\bar{y}_m = -\frac{\Delta'_m}{\Delta'} \qquad (m = 1, 2, \ldots, k) \tag{33}$$

$$\bar{y}_{k+1} = 1; \quad \bar{y}_{k+2} = \bar{y}_{k+3} = \ldots = \bar{y}_n = 0,$$

where

$$\Delta' = \begin{vmatrix} a_{11}, a_{21}, \ldots, a_{k1} \\ a_{12}, a_{22}, \ldots, a_{k2} \\ \cdots\cdots\cdots \\ a_{1k}, a_{2k}, \ldots, a_{kk} \end{vmatrix} \neq 0,$$

and Δ'_m is found from Δ' by substituting $a_{k+1,1} \ldots, a_{k+1,k}$ for the elements of the mth column. We write the condition for vector **b** with components $(b_1, \ldots, b_n)$ to be perpendicular to the **y** which we have obtained just now by solving system (32):

$$(\mathbf{b}, \mathbf{y}) = -\sum_{m=1}^{k} \frac{\Delta'_m}{\Delta'} b_m + b_{k+1} = 0$$

or

$$-\sum_{m=1}^{k} \Delta'_m b_m + \Delta' b_{k+1} = 0. \tag{34}$$

On interchanging rows and columns in the determinant Δ'_m then moving the mth row to the final position with the aid of $(k - m)$ interchanges of subsequent rows, we get:

$$-\Delta'_m = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, a_{1k} \\ \cdots & \cdots & \cdots \\ a_{m-1,1}, & a_{m-1,2}, & \ldots, a_{m-1,k} \\ a_{m+1,1}, & a_{m+1,2}, & \ldots, a_{m+1,k} \\ \cdots & \cdots & \cdots \\ a_{k1}, & a_{k2}, & \ldots, a_{kk} \\ a_{k+1,1}, & a_{k+1,2}, & \ldots, a_{k+1,k} \end{vmatrix} \cdot \quad (-1)^{k+1+m}.$$

This is precisely the cofactor of the element b_m in the characteristic determinant:

$$\Delta_{k+1} = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, a_{1k}, & b_1 \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1}, & a_{k2}, & \ldots, a_{kk}, & b_k \\ a_{k+1,1}, & a_{k+1,2}, & \ldots, b_{k+1,k}, & b_{k+1} \end{vmatrix},$$

so that condition (34) in fact expresses the vanishing of the characteristic determinant. Similarly, with $y_{k+s} = 1$, we get the condition

$\Delta_{k+s} = 0$. We thus arrive at the following result: *if the determinant of system* (30) *vanishes, the necessary and sufficient condition for the system to have a solution is that the vector* $(b_1, \ldots, b_n)$ *should be orthogonal to all the vectors yielding solutions of the homogeneous adjoint system* (32).

The general solution of system (30) is the sum of any particular solution of the system and the general soution of the corresponding homogeneous system obtained by replacing all the b_j in (30) by zeros. The general solution of the homogeneous system will contain $(n - k)$ arbitrary constants.

A further geometrical interpretation of the basic theorem regarding the solution of systems may be pointed out, since it is of importance later. Let us take n linear forms with n independent variables:

$$y_1 = a_{11} x_1 + \cdots + a_{1n} x_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$y_n = a_{n1} x_1 + \cdots + a_{nn} x_n.$$

We shall suppose that the x_s can take any complex values, $(y_1, \ldots, y_n)$ being regarded as the components of a certain vector. If the determinant $|\, a_{ik} \,|$ is not zero, we get definite values of x_k for any given y_k, and the previous formulae yield the whole of the n-dimensional space y. Now let the matrix $\|\, a_{ik} \,\|$ have rank $r < n$. We can assume without loss of generality that the rth order determinant at the top left corner is not zero. With this, the basic theorem regarding the solution of systems tells us the following: the set of values $(y_1, \ldots, y_n)$, obtained in accordance with the previous formulae, possesses the property that the values $y_1, \ldots, y_r$ may be chosen arbitrarily, yet once these are fixed, the remaining $y_{r+1}, \ldots, y_n$ are fully defined, being obtained from the vanishing condition for the characteristic determinants. This means in geometrical language that the previous formulae yield an r-dimensional subspace, formed by the vectors that are obtained by putting one of the $y_s (s = 1, 2, \ldots, r)$ equal to unity and the remainder equal to zero. All in all, then, if the rank of matrix $\|\, a_{ik} \,\|$ is r, the previous formulae yield a set of values $(y_1, \ldots, y_n)$ defining an r-dimensional subspace.

We have taken the case when the number of linear forms is equal to the number of variables x_s. We have in the general case:

$$y_1 = a_{11} x_1 + \cdots + a_{1n} x_n,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$y_m = a_{m1} x_1 + \cdots + a_{mn} x_n.$$

With arbitrary x_s, these formulae now define a subspace in the m-dimensional space, the number of dimensions of the subspace being equal to the rank of $\| a_{ik} \|$. The proof is the same as above.

16. Gram's determinant. Hadamard's inequality. Let us take m vectors:

$$\mathbf{x}^{(s)}\left(\mathbf{x}_1^{(s)}, \mathbf{x}_2^{(s)}, \ldots, \mathbf{x}_n^{(s)}\right) \quad (s = 1, 2, \ldots, m).$$

We form the mth order determinant of the scalar products $(\mathbf{x}^{(i)}, \mathbf{x}^{(k)})$ and introduce the special notation:

$$G\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}\right) = \left| \left(\mathbf{x}^{(i)}, \mathbf{x}^{(k)}\right) \right|_1^m =$$

$$= \begin{vmatrix} (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}), (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), \ldots, (\mathbf{x}^{(1)}, \mathbf{x}^{(m)}) \\ (\mathbf{x}^{(2)}, \mathbf{x}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}), \ldots, (\mathbf{x}^{(2)}, \mathbf{x}^{(m)}) \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ (\mathbf{x}^{(m)}, \mathbf{x}^{(1)}), (\mathbf{x}^{(m)}, \mathbf{x}^{(2)}), \ldots, (\mathbf{x}^{(m)}, \mathbf{x}^{(m)}) \end{vmatrix}. \tag{35}$$

This is known as *the Gram determinant of vectors*

$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}.$$

We distinguish the cases

$$m = n, \ m < n \ \text{and} \ m > n.$$

The general term of the Gram determinant has the form

$$(\mathbf{x}^{(i)}, \mathbf{x}^{(k)}) = \sum_{s=1}^{n} \mathbf{x}_s^{(i)} \overline{\mathbf{x}_s^{(k)}}.$$

With $m = n$, determinant (35) is equal to the product of the determinants:

$$\begin{vmatrix} \mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \ldots, \mathbf{x}_n^{(1)} \\ \mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \ldots, \mathbf{x}_n^{(2)} \\ \dots\dots\dots\dots \\ \mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}, \ldots, \mathbf{x}_n^{(n)} \end{vmatrix} \cdot \begin{vmatrix} \overline{\mathbf{x}_1^{(1)}}, \overline{\mathbf{x}_1^{(2)}}, \ldots, \overline{\mathbf{x}_1^{(n)}} \\ \overline{\mathbf{x}_2^{(1)}}, \overline{\mathbf{x}_2^{(2)}}, \ldots, \overline{\mathbf{x}_2^{(n)}} \\ \dots\dots\dots\dots \\ \overline{\mathbf{x}_n^{(1)}}, \overline{\mathbf{x}_n^{(2)}}, \ldots, \overline{\mathbf{x}_n^{(n)}} \end{vmatrix},$$

the multiplication rule of rows by columns being used. On noticing that the determinants are unchanged in value on interchanging rows with columns, we can say that the second factor is the complex conjugate of the first, so that with $m = n$ the Gram determinant (35) is equal to the square of the modulus of the determinant $| \mathbf{x}_k^{(i)} |_1^n$, formed by the components $\mathbf{x}_k^{(i)}$ of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Hence determinant (35) is positive if the vectors are linearly independent, and zero if they are linearly dependent [12]. With $m \neq n$, we have two rectangular matrices:

$$\begin{Vmatrix} \mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \ldots, \mathbf{x}_n^{(1)} \\ \mathbf{x}_1^{(2)}, \mathbf{x}_2^{(2)}, \ldots, \mathbf{x}_n^{(2)} \\ \dots\dots\dots\dots \\ \mathbf{x}_1^{(m)}, \mathbf{x}_2^{(m)}, \ldots, \mathbf{x}_n^{(m)} \end{Vmatrix} \ (36_1) \ \text{and} \ \begin{Vmatrix} \overline{\mathbf{x}_1^{(1)}}, \overline{\mathbf{x}_1^{(2)}}, \ldots, \overline{\mathbf{x}_1^{(m)}} \\ \overline{\mathbf{x}_2^{(1)}}, \overline{\mathbf{x}_2^{(2)}}, \ldots, \overline{\mathbf{x}_2^{(m)}} \\ \dots\dots\dots\dots \\ \overline{\mathbf{x}_m^{(1)}}, \overline{\mathbf{x}_m^{(2)}}, \ldots, \overline{\mathbf{x}_m^{(m)}} \end{Vmatrix}, \tag{36$_2$}$$

and the matrix corresponding to determinant (35) is the product of these two last matrices [7]. By the theorem proved in [7], determinant (35) vanishes for $m > n$. In this case, however, vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$ are linearly dependent [12]. With $m < n$, we have by the theorem proved above:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}) = \sum_{r_1<r_2<\ldots<r_m} X\begin{pmatrix} 1, 2, \ldots, m \\ r_1, r_2, \ldots, r_m \end{pmatrix} Y\begin{pmatrix} r_1, r_2, \ldots, r_m \\ 1, 2, \ldots, m \end{pmatrix},$$

where $X\begin{pmatrix} 1, & 2, & \ldots, & m \\ r_1, & r_2, & \ldots, & r_m \end{pmatrix}$ denotes a minor of array (36_1), and $Y\begin{pmatrix} r_1, & r_2, & \ldots, & r_m \\ 1, & 2, & \ldots, & m \end{pmatrix}$ a minor of (36_2). As above, the Y here is the conjugate of the X, and the last equation can be written:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}) = \sum_{r_1<r_2<\ldots<r_m} \left| X\begin{pmatrix} 1, 2, \ldots, m \\ r_1, r^2, \ldots, r_m \end{pmatrix} \right|^2. \qquad (37)$$

If $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$ are linearly independent, the rank of matrix (36_1) is equal to m [12], and at least one of the non-negative terms on the right-hand side of equation (37) is positive. If the vectors are linearly dependent, on the other hand, the rank of matrix (36_1) is less than m, all the mth order determinants appearing in the matrix vanish, and (37) implies that $G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, x^{(m)}) = 0$. Thus all three cases: $m = n$, $m > n$ and $m < n$, lead us to the following general theorem:

THEOREM. *The Gram determinant* $G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)})$ *is positive if the vectors* $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$ *are linearly independent, and zero if they are linearly dependent.*

We now prove a further formula for the Gram determinant. As a preliminary, we decide on the following notation. Let $\mathbf{x}$ be any vector of R_n and let the expansion be valid: $\mathbf{x} = \mathbf{y} + \mathbf{z}$, where $\mathbf{y}$ belongs to the subspace defined by vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$, and $\mathbf{z}$ is perpendicular to this subspace. We want to prove that

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{x}) = \|z\|^2 G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}). \qquad (38)$$

On taking into account the equations

$$(\mathbf{x}^{(s)}, \mathbf{x}) = (\mathbf{x}^{(s)}, \mathbf{y});\ (\mathbf{x}, \mathbf{x}^{(s)}) = (\mathbf{y}, \mathbf{x}^{(s)})$$

which follow from the orthogonality of $\mathbf{x}$ to all $\mathbf{x}^{(s)}$, and the equation $(\mathbf{x}, \mathbf{x}) = (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z})$ [13], we can write:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{x}) =$$

$$= \begin{vmatrix} (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(1)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(1)}, \mathbf{y}) \\ (\mathbf{x}^{(2)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(2)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(2)}, \mathbf{y}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (\mathbf{x}^{(m)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(m)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(m)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(m)}, \mathbf{y}) \\ (\mathbf{y}, \mathbf{x}^{(1)}), & (\mathbf{y}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{y}, \mathbf{x}^{(m)}), & (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z}) \end{vmatrix}.$$

On writing the elements of the last row as

$$(\mathbf{y}, \mathbf{x}^{(1)}) + 0, (\mathbf{y}, \mathbf{x}^{(2)}) + 0, \ldots, (\mathbf{y}, \mathbf{x}^{(m)}) + 0, (\mathbf{y}, \mathbf{y}) + (\mathbf{z}, \mathbf{z})$$

then expressing the determinant as the sum of two determinants in accordance with property IV of [3], we get

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{x}) = G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{y}) + \tag{39}$$
$$+\begin{vmatrix} (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(1)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(1)}, \mathbf{y}) \\ (\mathbf{x}^{(2)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(2)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(2)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(2)}, \mathbf{y}) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (\mathbf{x}^{(m)}, \mathbf{x}^{(1)}), & (\mathbf{x}^{(m)}, \mathbf{x}^{(2)}), & \ldots, & (\mathbf{x}^{(m)}, \mathbf{x}^{(m)}), & (\mathbf{x}^{(m)}, \mathbf{y}) \\ 0, & 0, & , \ldots, & 0, & \|\mathbf{z}\|^2 \end{vmatrix}.$$

The vector $\mathbf{y}$ belongs to the subspace defined by $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ and is therefore linearly expressible in terms of the $\mathbf{x}^{(S)}$; thus, by the theorem just proved:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{y}) = 0.$$

On expanding the determinant of (39) by the last row, we in fact get (38). The inequality:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{x}) \leqslant \|\mathbf{x}\|^2 G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}). \tag{40}$$

is an obvious consequence of (38). It may be remarked that, if the $x^{(S)}$ are linearly dependent, we have

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}, \mathbf{x}) = G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}) = 0.$$

If the $\mathbf{x}^{(S)}$ are linearly independent, the sign of equality is obtained in (40) when and only when $\mathbf{y} = 0$, i.e. when $\mathbf{x}$ is orthogonal to all the $\mathbf{x}^{(S)}$.

Repeated application of inequality (40) to the original Gram determinant $G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)})$ gives us

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}) \leqslant \|\mathbf{x}^{(1)}\|^2 \|\mathbf{x}^{(2)}\|^2 \ldots \|\mathbf{x}^{(m)}\|^2. \tag{41}$$

It must be borne in mind here that $G(\mathbf{x}^{(1)}) = \|\mathbf{x}^{(1)}\|^2$.

We have the sign of equality in (41) when and only when any two of the vectors are orthogonal (on the assumption that none is the null vector). This inequality leads easily to an inequality applicable to any determinant. Let Δ be an nth order determinant with elements a_{ik}. We shall look on the elements of the ith row as the components $(a_{i1}, a_{i2}, \ldots, a_{in})$ of a vector $\mathbf{x}^{(i)}$ of R_n. We form a new determinant with $\bar{a}_{ik}$, the conjugate elements to the a_{ik}, this determinant being obviously equal to $\bar{\Delta}$. The product of Δ and $\bar{\Delta}$, multiplying rows by rows, is the Gram determinant $G(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)})$, its value being equal to $\Delta\bar{\Delta}$, i.e. $|\Delta|^2$, by the theorem regarding the multiplication of determinants. Application of inequality (41) now leads us to Hadamard's inequality for the modulus of a determinant:

$$|\Delta|^2 \leqslant \sum_{k=1}^{n} |a_{1k}|^2 \cdot \sum_{k=1}^{n} |a_{2k}|^2 \ldots \sum_{k=1}^{n} |a_{nk}|^2. \tag{42}$$

If Δ has real elements, we can write:

$$\Delta^2 \leqslant \sum_{k=1}^{n} a_{1k}^2 \cdot \sum_{k=1}^{n} a_{2k}^2 \cdots \sum_{k=1}^{n} a_{nk}^2. \tag{43}$$

If the elements of the determinant satisfy

$$|a_{ik}| \leqslant M \qquad (i, k = 1, 2, \ldots, n),$$

it is obvious that

$$\sum_{k=1}^{n} |a_{ik}|^2 \leqslant nM^2,$$

and we now have from (42):

$$|\Delta| \leqslant n^{n/2} M^n. \tag{44}$$

It follows from our remarks above that the sign of equality is obtained in (42) when, and only when, any two of the vectors $\mathbf{x}^{(i)}$ are orthogonal.

We can obtain further inequalities for Gram determinants on the basis of a generalization of inequality (40).

Let X, Y, Z denote respectively sets of vectors of R_n. The generalization in question has the form:

$$G(X, Y, Z)\, G(X) \leqslant G(X, Z)\, G(X, Y). \tag{45}$$

This does not exclude the case of an empty set, i.e. one containing no vectors at all. If W is any such set, we have to take $G(W) = 1$.

On the basis of this inequality, we can write the following for Gram determinants:

$$G(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}) \leqslant \left[\prod_{k=1}^{m} G(\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k-1)}), \mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(m)}\right]^{1/(m-1)},$$

where Π is the product sign. Repeated application of this last expression leads to new expressions in which the Gram determinants contain a smaller number of vectors. In all these expressions, the sign of equality is obtained when, and only when, any two of the vectors are orthogonal. The expressions just given are due to M. K. Fage *(Dokl. Akad. Nauk SSSR,* 1946, **54**, No. 9).

17. Systems of linear differential equations with constant coefficients. We apply the results obtained to the problem of integrating systems of linear differential equations with constant coefficients. We take the system:

$$\left.\begin{aligned}
\mathbf{x}_1' &= a_{11}\mathbf{x}_1 + a_{12}\mathbf{x}_2 + \ldots + a_{1n}\mathbf{x}_n \\
\mathbf{x}_2' &= a_{21}\mathbf{x}_1 + a_{22}\mathbf{x}_2 + \ldots + a_{2n}\mathbf{x}_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\mathbf{x}_n' &= a_{n1}\mathbf{x}_1 + a_{n2}\mathbf{x}_2 + \ldots + a_{nn}\mathbf{x}_n
\end{aligned}\right\}, \tag{46}$$

where the x_j are required functions of t, the x_j' are their derivatives, and the a_{ik} are given constants. We shall seek a solution in the form:

$$x_1 = b_1 e^{\lambda t};\ x_2 = b_2 e^{\lambda t};\ \ldots;\ x_n = b_n e^{\lambda t}. \tag{47}$$

On substituting in system (46) and cancelling the factor $e^{\lambda t}$, we get a system of equations defining the constants $b_1, \ldots, b_n$:

$$\left.\begin{array}{l}(a_{11}-\lambda)\, b_1 + a_{12} b_2 + \ldots + a_{1n} b_n = 0\\ a_{21} b_1 + (a_{22}-\lambda)\, b_2 + \ldots + a_{2n} b_n = 0\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ a_{n1} b_1 + a_{n2} b_2 + \ldots + (a_{nn}-\lambda) b_n = 0.\end{array}\right\} \tag{48}$$

Since a non-trivial solution is required for the unknowns b_j, the determinant of this latter system must vanish, i.e. we have an equation of the form

$$\begin{vmatrix} a_{11}-\lambda, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22}-\lambda, & \ldots, & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn}-\lambda \end{vmatrix} = 0. \tag{49}$$

for the constant λ.

An equation of this type is generally known as a *secular equation.* It is familiar in the study of unconstrained vibrating mechanical systems in the particular case when the matrix of coefficients a_{ik} is symmetrical, i.e. $a_{ik} = a_{ki}$, and all the coefficients are real; this is a matter that we shall discuss later in connection with small vibrations. For the present, we shall discuss the general case. Equation (49) is an algebraic equation of degree n with highest term $(-\lambda)^n$; if it has n different roots:

$$\lambda = \lambda_1;\ \ldots;\ \lambda = \lambda_n,$$

substitution of each root λ_j for λ in the coefficients of system (48) gives us n homogeneous equations for the corresponding $b_1, \ldots, b_n$ with a vanishing determinant, so that a non-trivial solution in fact exists. Hence we have, from (47), n linearly independent solutions of system (46), and a linear combination of these yields the general solution of the system. If secular equation (49) has multiple roots, the solution of the problem is more difficult: to each root of (49) of multiplicity k there must correspond k linearly independent solutions of system (46), one solution having in fact the form (47), whilst the remainder in general contain a polynomial in t as a further factor. It must be remarked that the possibility does occur here — as is not

the case with a single equation with constant coefficients [II, 40] — of more than one (perhaps every) solution corresponding to a multiple root having the form (47). We shall not stop to consider this matter in more detail because a different method will later be used for solving system (46), based on the theory of functions of a complex variable.

We return to secular equation (49) which is fundamental to our problem. The solution, even if approximate, of this equation presents practical difficulties for large n, due to the fact that the unknown λ appears along the diagonal and not in a single row or column. Expansion of the left-hand side in powers of λ requires a large number of computations, as indicated above in [4]. We shall describe a method of transformation of equation (49) to a form more convenient in practice, by means of which the unknown λ is brought into a single column. This method is due to Prof. A. N. Krilov, who gave the first exposition of it in his article "The numerical solution of equations determining the frequencies of small vibrations in material systems in engineering" *(Izv. Akad. Nauk SSSR*, 1931).

We form a linear combination of the required magnitudes:

$$\xi = a_{01} x_1 + a_{02} x_2 + \cdots + a_{0n} x_n, \tag{50}$$

where the a_{0j} are numerical coefficients chosen in any manner. We now differentiate equation (50) n times with respect to t, each time replacing the derivatives x'_j on the right-hand side by their expressions from system (46). We get the $(n + 1)$ equations:

$$\left.\begin{array}{l} \xi = a_{01} x_1 + a_{02} x_2 + \cdots + a_{0n} x_n \\ \xi' = a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \xi^{(n-1)} = a_{n-1,1} x_1 + a_{n-1,2} x_2 + \cdots + a_{n-1,n} x_n \\ \xi^{(n)} = a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n. \end{array}\right\} \tag{51}$$

Let the determinant formed from the coefficients a_{ik} appearing in the first n equations differ from zero. The first n equations then give us expressions for the x_j in terms of $\xi, \xi', \ldots, \xi^{(n-1)}$, and substitution of these expressions in the last equation gives us an nth order equation for ξ. Elimination of the x_j from the $(n + 1)$ equations (51) can be carried out directly with the aid of determinants. We first re-write these equations as

$$\xi x_0 + a_{01} x_1 + a_{02} x_2 + \cdots + a_{0n} x_n = 0,$$
$$\xi' x_0 + a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\xi^{(n)} x_0 + a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n = 0,$$

where $x_0 = -1$, then we consider these as a homogeneous system in the magnitudes

$$x_0, x_1, \ldots, x_n.$$

The determinant of this homogeneous system must vanish, and this in fact gives us the required result of elimination:

$$\begin{vmatrix} \xi, & a_{01}, a_{02}, \dots, a_{0n} \\ \xi', & a_{11}, a_{12}, \dots, a_{1n} \\ \dots & \dots \\ \xi^{(n)}, & a_{n1}, a_{n2}, \dots, a_{nn} \end{vmatrix} = 0. \tag{52}$$

We shall seek the solution of this equation in the form

$$\xi = e^{\lambda t}.$$

On substituting this in the first column of determinant (52), taking the factor $e^{\lambda t}$ from the column outside the determinant sign, then cancelling it,we get the following equation for λ:

$$\begin{vmatrix} 1, & a_{01}, a_{02}, \dots, a_{0n} \\ \lambda, & a_{11}, a_{12}, \dots, a_{1n} \\ \dots & \dots \\ \lambda^n, & a_{n1}, a_{n2}, \dots, a_{nn} \end{vmatrix} = 0. \tag{53}$$

It may easily be shown that, given our assumption, equation (53) has the same roots as (49). For, let $\lambda = \lambda_0$ be a solution of (53); then we have a solution of (52) of the form:

$$\xi = Ce^{\lambda_0 t}, \tag{54}$$

where C is an arbitrary constant. The first n equations of system (51) now give us solutions of type (47) for the x_j with $\lambda = \lambda_0$, i.e. $\lambda = \lambda_0$ is in fact a root of equation (49). Conversely, if $\lambda = \lambda_0$ is a root of (49), we have a solution of type (47) for the x_j with $\lambda = \lambda_0$, where the b_j are numerical constants, not all of which are zero. On substituting these expressions for the x_j in the first of equations (51), we in fact obtain a solution for ξ of type (54), this solution being certainly non-zero, since otherwise we should have

$$\xi = \xi' = \dots = \xi^{(n-1)} = 0,$$

whence it would immediately follow from the first n equations of system(51) that

$$x_1 = x_2 = \dots = x_n = 0.$$

Thus every root $\lambda = \lambda_0$ of equation (49) is in fact a root of equation (53). We have now shown that, given our assumption, equation (53) has the same roots as (49). Applications of this method to numerical examples, together with a discussion of the case when our assumption no longer holds, may be found in the article by Prof. Krilov quoted above.

Simpler working is obtained if formula (50) is taken as $\xi = x_1$. In this case, (53) becomes

$$\begin{vmatrix} 1, & 1, & 0, \dots, 0 \\ \lambda, & a_{11}, & a_{12}, \dots, a_{1n} \\ \dots & \dots & \dots \\ \lambda^n, & a_{n1}, & a_{n2}, \dots, a_{nn} \end{vmatrix} = 0.\dagger$$

† A. Danilevskii has proposed a neat method for transforming the secular determinant in *Mat. Sbornik*, 2, Sec. 1.

We consider, instead of (46), the system of second order equations:

$$\left.\begin{array}{l} x_1'' = a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n \\ x_2'' = a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_n'' = a_{n1} x_1 + a_{n2} x_2 + \ldots + a_{nn} x_n . \end{array}\right\} \qquad (55)$$

Systems of this type are often encountered in mechanics. If we seek a solution in the form

$$x_j = b_j \cos(\lambda t + \varphi),$$

we obtain an equation for λ of the form:

$$\begin{vmatrix} a_{11} + \lambda^2, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22} + \lambda^2, & \ldots, & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} + \lambda^2 \end{vmatrix} = 0 . \qquad (56)$$

The constants b_j are defined by a system analogous to system (48), and φ remains arbitrary.

Finally, with systems such as

$$\left.\begin{array}{l} x_1'' = a_{11} x_1 + \ldots + a_{1n} x_n + c_{11} x_1' + \ldots + c_{1n} x_n' \\ x_2'' = a_{21} x_1 + \ldots + a_{2n} x_n + c_{21} x_1' + \ldots + c_{2n} x_n' \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n'' = a_{n1} x_1 + \ldots + a_{nn} x_n + c_{n1} x_1' + \ldots + c_{nn} x_n' \end{array}\right\} \qquad (57)$$

which also include the first derivatives, we again seek a solution of type (47), and arrive at a secular equation of the form:

$$\begin{vmatrix} a_{11} + c_{11}\lambda - \lambda^2, & a_{12} + c_{12}\lambda & , \ldots, & a_{1n} + c_{1n}\lambda \\ a_{21} + c_{21}\lambda, & a_{22} + c_{22}\lambda - \lambda^2, & \ldots, & a_{2n} + c_{2n}\lambda \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} + c_{n1}\lambda, & a_{n2} + c_{n2}\lambda & , \ldots, & a_{nn} + c_{nn}\lambda - \lambda^2 \end{vmatrix} = 0. \qquad (58)$$

If we introduce the supplementary unknowns:

$$x_{n+1} = x_1'; \; x_{n+2} = x_2'; \; \ldots; \; x_{2n} = x_n', \qquad (59)$$

we can reduce system (57) to $2n$ first order equations, n of these being obtained from (57) by substituting

$$x_j'' = x_{n+j}' \text{ and } x_j' = x_{n+j} \qquad (j = 1, 2, \ldots, n),$$

whilst the remaining n are equations (59).

18. Functional determinants. Let us take n functions of n variables:

$$\varphi_1(x_1, x_2, \ldots, x_n); \; \varphi_2(x_1, x_2, \ldots, x_n); \ldots; \; \varphi_n(x_1, x_2, \ldots, x_n). \quad (60)$$

The *functional determinant* of these functions in the variables x_s is the nth order determinant whose elements are given by $a_{ik} = \partial\varphi_i/\partial x_k$.

We bring in the special notation for the functional determinant:

$$\frac{D(\varphi_1, \ldots, \varphi_n)}{D(x_1, \ldots, x_n)} = \begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1}, & \frac{\partial \varphi_1}{\partial x_2}, & \ldots, & \frac{\partial \varphi_1}{\partial x_n} \\ \frac{\partial \varphi_2}{\partial x_1}, & \frac{\partial \varphi_2}{\partial x_2}, & \ldots, & \frac{\partial \varphi_2}{\partial x_n} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial \varphi_n}{\partial x_1}, & \frac{\partial \varphi_n}{\partial x_2}, & \ldots, & \frac{\partial \varphi_n}{\partial x_n} \end{vmatrix} . \tag{61}$$

We have already encountered determinants of this type in the change of variables in multiple integrals [II, 57 and 60]. If we have the change of variables on a plane:

$$x = \varphi(u, v); \quad y = \psi(u, v) , \tag{62}$$

where the point (u, v) becomes the point (x, y), the absolute value of the functional determinant (Jacobian)

$$\frac{D(\varphi, \psi)}{D(u, v)} \tag{63}$$

gives the coefficient of change of area at the point (u, v) under transformation (62), on the assumption that the partial derivatives of the functions of (62) with respect to u and v are continuous, and that determinant (63) does not vanish, in the domain over which the transformation is applied. Similarly, if we have the point transformation in three-dimensional space:

$$x = \varphi(q_1, q_2, q_3); \quad y = \psi(q_1, q_2, q_3); \quad z = \omega(q_1, q_1, q_3) ,$$

where the point with coordinates (q_1, q_2, q_3) becomes the point (x, y, z) and the volume (V_1) becomes volume (V), the formula for change of variables in the triple integral may be written [II, 60]:

$$\iiint\limits_{V} f(x, y, z)\, dx\, dy\, dz = \iiint\limits_{(V_1)} f[\varphi, \psi, \omega] \; |D|\, dq_1\, dq_2\, dq_3,$$

where

$$D = \frac{D(\varphi, \psi, \omega)}{D(q_1, q_2, q_3)} ,$$

and $|D|$ is the coefficient of cubical change at a given point on transforming from (q_1, q_2, q_3) to (x, y, z).

We might have considered the single function of a single independent variable:

$$u = f(x)$$

in exactly the same way, as a transformation of points on the axis OX, in which the point abscissa x takes up a new position with abscissa u. The absolute value $|f'(x)|$ of the derivative obviously characterizes the change in linear measure at a given point. Everything that has been said may be extended to point transformations in n-dimensional space and to the change of variables in n-tuple integrals [II, 98].

Having explained the two and three-dimensional analogies between functional determinants and derivatives, we now show that there are analogies as regards their formal properties.

Let us take the system of functions

$$\varphi_1(y_1, \ldots, y_n), \ldots, \varphi_n(y_1, \ldots, y_n),$$

and instead of $y_1, \ldots, y_n$ being independent variables, let them be in turn functions of $x_1, \ldots, x_n$, so that in the last analysis the φ_i are functions of the x_i. We can form three functional determinants:

$$\frac{D(\varphi_1, \ldots, \varphi_n)}{D(y_1, \ldots, y_n)}; \quad \frac{D(\varphi_1, \ldots, \varphi_n)}{D(x_1, \ldots, x_n)}; \quad \frac{D(y_1, \ldots, y_n)}{D(x_1, \ldots, x_n)}.$$

The elements of these determinants are respectively

$$\frac{\partial\varphi_i}{\partial y_k}; \quad \frac{\partial\varphi_i}{\partial x_k}; \quad \frac{\partial y_i}{\partial x_k}.$$

But we have by the rule for differentiating functions of a function:

$$\frac{\partial\varphi_i}{\partial x_k} = \frac{\partial\varphi_i}{\partial y_1}\cdot\frac{\partial y_1}{\partial x_k} + \ldots + \frac{\partial\varphi_i}{\partial y_n}\cdot\frac{\partial y_n}{\partial x_k},$$

and the determinant multiplication rule of rows by columns gives us an equation expressing the first property of functional determinants:

$$\frac{D(\varphi_1, \ldots, \varphi_n)}{D(x_1, \ldots, x_n)} = \frac{D(\varphi_1, \ldots, \varphi_n)}{D(y_1, \ldots, y_n)}\cdot\frac{D(y_1, \ldots, y_n)}{D(x_1, \ldots, x_n)}. \tag{64}$$

This is analogous to the rule for differentiating functions of a function with a single independent variable.

A further property of functional determinants is as follows. The system of functions φ_i can be considered as a transformation of variables x_i to the new variables φ_i:

$$\varphi_i = \varphi_i(x_1, \ldots, x_n) \quad (i = 1, 2, \ldots, n). \tag{65}$$

We first notice the particular case of the so-called identity transformation:

$$\varphi_1 = x_1; \quad \varphi_2 = x_2; \ldots; \quad \varphi_n = x_n.$$

Its functional determinant is

$$\begin{vmatrix} 1, & 0, & 0, & \ldots, & 0 \\ 0, & 1, & 0, & \ldots, & 0 \\ 0, & 0, & 1, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \ldots, & 1 \end{vmatrix} = 1 .$$

We shall imagine that equations (65) have been solved with respect to the x_i, so that the x_i can be written in terms of the φ_i:

$$x_i = x_i(\varphi_1, \ldots, \varphi_n) \quad (i = 1, 2, \ldots, n) . \tag{66}$$

Transformation (66) is naturally known as the inverse of (65). If expressions (66) are substituted in the right-hand sides of (65), we get the identities: $\varphi_1 = \varphi_1; \ldots, \varphi_n = \varphi_n$, or in other words, we get the identity transformation. On now applying formula (64) to this particular case, we have to put $y_i = x_i$ and $x_i = \varphi_i$, whilst we have the functional determinant of the identity transformation on the left-hand side:

$$\frac{D(\varphi_1, \ldots, \varphi_n)}{D(\varphi_1, \ldots, \varphi_n)} = \frac{D(\varphi_1, \ldots, \varphi_n)}{D(x_1, \ldots, x_n)} \cdot \frac{D(x_1, \ldots, x_n)}{D(\varphi_1, \ldots, \varphi_n)}$$

or

$$\frac{D(\varphi_1, \ldots, \varphi_n)}{D(x_1, \ldots, x_n)} \cdot \frac{D(x_1, \ldots, x_n)}{D(\varphi_1, \ldots, \varphi_n)} = 1 , \tag{67}$$

i.e. *the product of the functional determinants of the direct and inverse transformations is unity.* This is analogous to the property of the derivatives of inverse functions in the case of a single independent variable.

We now explain the meaning of the condition that the functional determinant

$$\frac{D(\varphi_1, \varphi_2, \ldots, \varphi_n)}{D(x_1, x_2, \ldots, x_n)} \tag{68}$$

of the functions

$$\varphi_1(x_1, \ldots, x_n);\ \varphi_2(x_1, \ldots, x_n);\ \ldots;\ \varphi_n(x_1, \ldots, x_n)$$

with respect to the variables x_s is identically equal to zero. Suppose that these functions are connected by the functional relationship

$$F(\varphi_1 \ldots, \varphi_n) = 0 , \tag{69}$$

this equation being an identity in the independent variables x_s. If we differentiate with respect to all the independent variables, we get the n identities:

$$\left.\begin{aligned} &\frac{\partial F}{\partial \varphi_1} \cdot \frac{\partial \varphi_1}{\partial x_1} + \ldots + \frac{\partial F}{\partial \varphi_n} \cdot \frac{\partial \varphi_n}{\partial x_1} = 0 \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ &\frac{\partial F}{\partial \varphi_1} \cdot \frac{\partial \varphi_1}{\partial x_n} + \ldots + \frac{\partial F}{\partial \varphi_n} \cdot \frac{\partial \varphi_n}{\partial x_n} = 0 . \end{aligned}\right\} \tag{70}$$

We can look on these n identities as linear equations in the n quantities

$$\frac{\partial F}{\partial \varphi_1}, \ldots, \frac{\partial F}{\partial \varphi_n},$$

where it is clear that the quantities cannot vanish identically at the same time, since otherwise F would contain none of the φ_i. The determinant of homogeneous system (70) must therefore vanish, which amounts to the vanishing of functional determinant (68). The presence of functional relationship (69) thus implies that functional determinant (68) vanishes identically. We shall not dwell on the proof of the converse, which is also true, i.e. *the vanishing identically of functional determinant* (68) *is the necessary and sufficient condition for a relationship to exist between the functions* $\varphi_i(x_1, \ldots, x_n)$.†

We take the example of three functions of three independent variables:

$$\varphi_1 = x_1^2 + x_2^2 + x_3^2; \quad \varphi_2 = x_1 + x_2 + x_3; \quad \varphi_3 = x_1 x_2 + x_1 x_3 + x_2 x_3 . \tag{71}$$

It may easily be verified that the following relationship exists between these:

$$\varphi_2^2 - \varphi_1 - 2\varphi_3 = 0 .$$

We form the functional determinant for functions (71):

$$\frac{D(\varphi_1, \varphi_2, \varphi_3)}{D(x_1, x_2, x_3)} = \begin{vmatrix} 2x_1, & 2x_2, & 2x_3 \\ 1, & 1, & 1 \\ x_2 + x_3, & x_1 + x_3, & x_1 + x_2 \end{vmatrix} .$$

We suggest that the reader show that this determinant is identically zero.

19. Implicit functions. We proved the existence theorem for the implicit function defined by a single equation in Vol. I [I, 159]. We now generalize this for the case of a system of equations. The original theorem will first be re-stated: let $x = x_0$, $y = y_0$ be a solution of the equation

$$F(x, y) = 0 \tag{72}$$

and let $F(x, y)$ and its first order partial derivatives be continuous at and in the neighbourhood of $x = x_0$, $y = y_0$; also, let the partial derivative $F'(x, y)$ differ from zero at $x = x_0$, $y = y_0$. For x sufficiently close to x_0, equation (72) now defines a unique function $y(x)$ which is continuous, has a derivative, and satisfies the condition $y(x_0) = y_0$. As we have already mentioned, it can similarly be shown that the equation

$$F(x, y, z) = 0 ,$$

having the solution $x = x_0$, $y = y_0$, $z = z_0$, where $F(x, y, z)$ and its first order partial derivatives are continuous in the neighbourhood of this solution, and $F'_z(x_0, y_0, z_0) \neq 0$, uniquely defines a function $z(x, y)$ which is continuous in the neighbourhood of $x = x_0$, $y = y_0$, possesses derivatives with respect to x and y, and satisfies the condition $z(x_0, y_0) = z_0$. We now consider the system of two equations:

$$\varphi(x, y, z) = 0; \quad \psi(x, y, z) = 0 . \tag{73}$$

† It must be pointed out that our discussion regarding system (70) is of a formal nature and is not, strictly speaking, a proof.

Let this system have a solution $x = x_0$, $y = y_0$, $z = z_0$, let $\varphi(x, y, z)$, $\psi(x, y, z)$ and their partial derivatives be continuous in the neighbourhood of the solution, and let the functional determinant

$$\frac{D(\varphi, \psi)}{D(y, z)} = \begin{vmatrix} \frac{\partial \varphi}{\partial y}, & \frac{\partial \varphi}{\partial z} \\ \frac{\partial \psi}{\partial y}, & \frac{\partial \psi}{\partial z} \end{vmatrix} = \frac{\partial \varphi}{\partial y}\,\frac{\partial \psi}{\partial z} - \frac{\partial \varphi}{\partial z}\,\frac{\partial \psi}{\partial y} \tag{74}$$

be non-zero at the solution. With these conditions, and with x sufficiently close to x_0, system (73) *defines a unique system of functions $y(x)$, $z(x)$ which are continuous, have first order derivatives, and satisfy the condition $y(x_0) = y_0$, $z(x_0) = z_0$.*

Since expression (74) differs from zero at $x = x_0$, $y = y_0$, $z = z_0$, at least one of the partial derivatives $\partial\psi/\partial y$ or $\partial\psi/\partial z$ must differ from zero. Suppose that say $\partial\psi/\partial z$ is non-zero at the solution. By the theorem stated above, the second of equations (73) uniquely defines a function $z(x, y)$. On substituting this function in the first equation of the system, we get an equation in the variables x and y:

$$\varphi[x, y, z(x, y)] = 0 . \tag{75}$$

It only remains for us to show, in order to prove the theorem, that the partial derivative with respect to y of the left-hand side of (75) differs from zero for $x = x_0$, $y = y_0$. This partial derivative is given by

$$\left(\frac{\partial \varphi}{\partial y}\right) = \frac{\partial \varphi}{\partial y} + \frac{\partial \varphi}{\partial z}\cdot\frac{\partial z}{\partial y}, \tag{76}$$

where $(\partial\varphi/\partial y)$ is the total derivative of $\varphi(x, y, z)$ with respect to the argument y. Since $z(x, y)$ is the solution of the second of equations (73), we have the identity:

$$\psi\,[x, y, z\,(x, y)] = 0 .$$

We differentiate this identity with respect to y:

$$\frac{\partial \psi}{\partial y} + \frac{\partial \psi}{\partial z}\cdot\frac{\partial z}{\partial y} = 0. \tag{77}$$

We multiply both sides of (76) by $\partial\psi/\partial z$ and add to (77) after multiplying both sides of (77) by $-\partial\varphi/\partial z$. This gives us after simple working:

$$\frac{\partial \psi}{\partial z}\cdot\left(\frac{\partial \varphi}{\partial y}\right) = \frac{D\,(\varphi, \psi)}{D\,(y, z)} .$$

The function $z(x, y)$ becomes z_0 with $x = x_0$, $y = y_0$, and $\partial\psi/\partial z$ and (74) differ from zero with these values of the variables, so that $(\partial\varphi/\partial y)$ is also non-zero. Consequently equation (75) defines a unique function $y(x)$. On substituting this in $z(x, y)$, we in fact obtain z as a function of x. This proof is possible with several independent variables instead of x.

The implicit function theorem may be stated as follows in the general case: *Let the system of equations*

$$F_1\,(x_1, \ldots, x_m, y_1, \ldots, y_n) = 0;\ \ldots;\ F_n(x_1, \ldots, x_m, y_1, \ldots, y_n) = 0, \tag{78}$$

have the solution

$$x_k = x_k^{(0)}, \quad y_l = y_l^{(0)} \qquad \begin{pmatrix} k = 1, \ldots, m \\ l = 1, \ldots, n \end{pmatrix}; \tag{79}$$

let the F_l *be continuous and have continuous first order partial derivatives in the neighbourhood of solution* (79), *and finally, let the functional determinant*

$$\frac{D(F_1 \ldots, F_n)}{D(y_1, \ldots, y_n)} = \begin{vmatrix} \frac{\partial F_1}{\partial y_1}, & \frac{\partial F_1}{\partial y_2}, & \ldots, & \frac{\partial F_1}{\partial y_n} \\ \frac{\partial F_2}{\partial y_1}, & \frac{\partial F_2}{\partial y_2}, & \ldots, & \frac{\partial F_2}{\partial y_n} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial F_n}{\partial y_1}, & \frac{\partial F_n}{\partial y_2}, & \ldots, & \frac{\partial F_n}{\partial y_n} \end{vmatrix} \tag{80}$$

differ from zero at solution (79). *Then for* x_k *sufficiently close to* $x_k^{(0)}$, *equations* (78) *define a unique system of functions* $y_l(x_1, \ldots, x_n)$ *that are continuous, possess first order derivatives, and satisfy the conditions* $y_l(x_1^{(0)}, \ldots, x_k^{(0)}) = y_l^{(0)}$.

We shall sketch out the proof of this theorem. We suppose it to be true for $(n-1)$ equations (it is in fact valid for $n = 1$ and $n = 2$) then show that it is true for n equations. By expanding determinant (80) by its first column, we can say that at least one of the corresponding cofactors must be non-zero for values (79), since (80) is itself non-zero at the solution by hypothesis. We can choose the subscripts for the F_l in such a way that the cofactor of $\partial F_1/\partial y_1$ is non-zero. This cofactor consists of the functional determinant of $F_2, \ldots, F_n$ with respect to variables $y_2, \ldots, y_n$. By the theorem for the system of $(n-1)$ equations, the equations

$$F_2(x_1, \ldots, x_m, y_1, \ldots, y_n) = 0; \ldots; F_n(x_1, \ldots, x_m, y_1 \ldots y_n) = 0 \tag{81}$$

uniquely define the functions

$$y_2 = \varphi_2(x_1, \ldots; x_m, y_1); \ \ldots; \ y_n = \varphi_n(x_1 \ldots, x_m, y_1). \tag{82}$$

On substituting these functions in the first of equations (78), we get the equation for y_1:

$$F_1(x_1, \ldots, x_m, y_1, \varphi_2, \ldots, \varphi_n) = 0. \tag{83}$$

It remains for us to verify that the total derivative of the left-hand side of this equation with respect to y_1 differs from zero for values (79). The derivative is given by

$$\left(\frac{\partial F_1}{\partial y_1}\right) = \frac{\partial F_1}{\partial y_1} + \sum_{s=2}^{n} \frac{\partial F_1}{\partial \varphi_s} \cdot \frac{\partial \varphi_s}{\partial y_1}. \tag{84}$$

On substituting functions (82) in the left-hand sides of equations (81), we obtain identities which we differentiate with respect to y_1:

$$\frac{\partial F_l}{\partial y_1} + \sum_{s=2}^{n} \frac{\partial F_l}{\partial \varphi_s} \cdot \frac{\partial \varphi_s}{\partial y_1} = 0 \quad (l = 2, \ldots, n). \tag{85}$$

Let $A_1, A_2, \ldots, A_n$ denote the cofactors of the elements of the first column

of (80). On multiplying (84) by A_1 and (85) by A_l then subtracting the latter equation from the former, we obtain the equation:

$$A_1\left(\frac{\partial F_1}{\partial y_1}\right) = \sum_{l=1}^{n} \frac{\partial F_l}{\partial y_1} A_l + \sum_{s=2}^{n}\left[\sum_{l=1}^{n} \frac{\partial F_l}{\partial \varphi_s} A_l\right] \frac{\partial \varphi_s}{\partial y_1}.$$

The first sum on the right-hand side yields determinant (80) which we shall write simply as D for brevity, whilst the summation over l in the second term represents the sum of products of elements belonging to a column other than the first of D with the cofactors of the corresponding elements of the first column, i.e. the sum is zero. We remark here that differentiation with respect to φ_s is exactly the same as differentiation with respect to y_s. The above equation thus reduces to

$$A_1\left(\frac{\partial F_1}{\partial y_1}\right) = D.$$

Since A_1 and D do not vanish at solution (79), the same can be said regarding the derivative with respect to y_1 of the left-hand side of equation (83); hence (83) yields a unique function $y_1(x_1, x_2, \ldots, x_m)$. Substitution in functions (82) gives us the final result.

The inversion theorem for systems of functions is a particular case of the implicit function theorem. *Let the equations be given:*

$$y_k = f_k(x_1 \ldots, x_n) \quad (k = 1, 2, \ldots, n). \tag{86}$$

Let the functions f_k and their first order derivatives be continuous in the neighbourhood of $x_k = x_k^{(0)}$ $(k = 1, 2, \ldots, n)$, for which values the functional determinant

$$\frac{D(f_1, \ldots, f_n)}{D(x_1, \ldots, x_n)} \tag{87}$$

is different from zero. Then equations (86) *uniquely define $x_k(y_1, \ldots, y_n)$ as functions of $y_1, \ldots, y_n$ in the neighbourhood of $y_k^{(0)} = f_k(x_1^{(0)}, \ldots, x_n^{(0)})$, these functions being continuous and having first order derivatives, whilst satisfying $x_k(y_1^{(0)}, \ldots, y_n^{(0)}) = x_k^{(0)}$.*

We prove this theorem simply by taking the equations

$$f_k(x_1, \ldots, x_n) - y_k = 0 \quad (k = 1, 2, \ldots, n)$$

and applying the implicit function theorem, the role of y_l being played by x_k.

If the f_k are linear homogeneous functions of the variables x_k, system (86) has the form:

$$y_k = a_{k1} x_1 + a_{k2} x_2 + \cdots + a_{kn} x_n.$$

Determinant (87) reduces in this case to the determinant $|a_{ik}|$ of the coefficients a_{ik}, and the existence of a unique solution of the system depends on Cramer's theorem.

CHAPTER II

LINEAR TRANSFORMATIONS AND QUADRATIC FORMS

20. Coordinate transformations in three-dimensional space. A linear transformation in n variables is defined by:

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n' &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{aligned}\right\} \qquad (1)$$

This can be interpreted as the passage from a vector $(x_1, \ldots, x_n)$ of n-dimensional space to another vector $(x_1', \ldots, x_n')$. Alternatively, we can regard $(x_1, \ldots, x_n)$ as the coordinates of a point in n-dimensional space and (1) as the passage from this point to another.

Yet another interpretation is possible: we can regard $(x_1, \ldots, x_n)$ and $(x_1', \ldots, x_n')$ as the components of the same vector (or coordinates of the same point), but with different choices of axes. Expressions (1) now give the transformation of components (coordinates) on passing from one coordinate system to the other. Expressions of type (1) with $n = 2$ and $n = 3$ have already been encountered a number of times.

The first part of the present chapter is devoted to a detailed study of linear transformation (1). We start with real three-dimensional space for the sake of greater clarity, then pass to the general case of complex n-dimensional space. Our discussion for three-dimensional space begins with the most elementary case, when (1) corresponds to passage from one set of rectangular axes to another. On measuring vectors from the origin, we can obviously take (x_1, x_2, x_3) as either the components of a vector or the coordinates of its terminus.

The expressions for transforming Cartesian coordinates are familiar from analytic geometry:

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned}\right\}, \qquad (2)$$

where the a_{ik} are the cosines of the angles formed by the new axes with the old and are given by the following table:

	X_1	X_2	X_3
X_1'	a_{11}	a_{12}	a_{13}
X_2'	a_{21}	a_{22}	a_{23}
X_3'	a_{31}	a_{32}	a_{33}

(3)

We know that the array of coefficients in (3) has the following properties: the sum of the squares of the elements of each row and column is equal to unity, and the sum of the products of corresponding elements of two different rows or columns is zero. The magnitude of the determinant

$$|a_{ik}|$$

is clearly equal [5] to the volume of the rectangular parallelepiped with unit sides directed along the new axes, i.e. it is unity if the axes have the same orientation, and (-1) if the orientation is different. The inverse transformation from (x_1', x_2', x_3') to (x_1, x_2, x_3) will clearly be:

$$\left.\begin{aligned} x_1 &= a_{11}x_1' + a_{21}x_2' + a_{31}x_3' \\ x_2 &= a_{12}x_1' + a_{22}x_2' + a_{32}x_3' \\ x_3 &= a_{13}x_1' + a_{23}x_2' + a_{33}x_3' \end{aligned}\right\} . \tag{2_1}$$

In other words, the inverse transformation to (2) is simply obtained by interchange of rows with columns in the array of coefficients of (2). The determinant of the inverse transformation is obviously equal to the determinant of (2).

We now show that the properties mentioned of the coefficients of (2) can be obtained by satisfying a single requirement that follows at once from the geometrical nature of our problem. We look for all the real transformations of type (2) such that

$$x_1'^2 + x_2'^2 + x_3'^2 = x_1^2 + x_2^2 + x_3^2. \tag{4}$$

This statement of the problem enables us to generalize our discussion of transformations to the case of space with any number of dimensions. What we do is show that the transformations required by the new problem are the same as those discussed above, i.e. we show that requirement (4) leads to the previous relationships between the a_{ik}. We substitute from (2) in the left-hand side of (4), remove the brackets,

then equate the coefficients of the squares of the variables to unity, and the products of different coefficients to zero; this gives us six relationships of the type:

$$a_{1k}a_{1l} + a_{2k}a_{2l} + a_{3k}a_{3l} = \delta_{kl} \qquad (k, l = 1, 2, 3), \tag{5}$$

where

$$\delta_{kl} = 0 \text{ for } k \neq l \text{ and } \delta_{kk} = 1, \tag{6}$$

i.e. the sum of the squares of the elements of each column is unity and the sum of the products of corresponding elements of different columns is zero. These conditions are generally known as *orthogonality conditions in regard to columns*. It now follows immediately that the elements of each column are the direction-cosines of a certain straight line, and that the straight lines corresponding to different columns are mutually perpendicular. This implies in turn that in the present case transformation (2) coincides with that considered above, and that we have orthogonality in regard to rows as well as to columns.

We can look on (2) as a transformation of space with fixed axes, instead of a coordinate transformation in a fixed space. Suppose first that the transformation determinant is equal to $(+1)$, i.e. both systems of axes have the same orientation. We can now rotate the space like a solid body about the origin together with the axes (X_1', X_2', X_3') so that these axes coincide with (X_1, X_2, X_3) which we take to be fixed during the rotation and to which we refer the coordinates of every point both before and after rotation. If a point M had the coordinates (x_1, x_2, x_3) before rotation, it takes up a new position M' as a result of rotation and has the new coordinates (x_1', x_2', x_3'). Since the point M moves with the axes (X_1', X_2', X_3'), the coordinates (x_1', x_2', x_3') of M' with respect to (X_1, X_2, X_3), with which (X_1', X_2', X_3') have come to coincide as a result of the rotation, will be the same as the coordinates of M with respect to (X_1', X_2', X_3') before rotation. Hence it may be seen that expressions (2) represent, in the case of the $(+1)$ determinant, a transformation of the coordinates of a point as a result of rotating the space.

Now let $|a_{ik}|$ be equal to (-1). We consider instead of (2) the transformation

$$x_i'' = -a_{i1}x_1 - a_{i2}x_2 - a_{i3}x_3 \qquad (i = 1, 2, 3).$$

Its coefficients possess properties (5) as before, whilst its determinant now has the value $(+1)$, i.e. it corresponds to a rotation of the space about the origin. In order to obtain the coordinates (x_1', x_2', x_3'), we

have to carry out the further transformation:

$$x'_1 = -x''_1; \quad x'_2 = -x''_2; \quad x'_3 = -x''_3$$

which is a symmetry transformation about the origin, inasmuch as the signs of all the coordinates are changed. Thus transformation (2) corresponds, in the case of the (-1) determinant, to a rotation of the space about the origin followed by a symmetrical shift with respect to the origin.

We saw above that the nine coefficients a_{ik} have to satisfy the six relationships (5). This means that they are expressible in terms of three independent parameters. We shall indicate one possible choice of parameters in the case of a rotation of space about the origin.

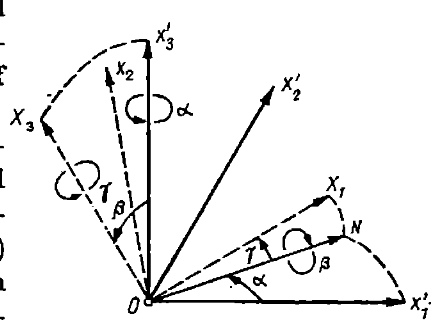

FIG. 1

We bring in two systems of coordinate axes: (X'_1, X'_2, X'_3) is a fixed system to which all the coordinates are referred, whilst (X_1, X_2, X_3) has an invariable relationship with the rotating space. In order to define the rotation, we have to establish three parameters defining the position of the second system of axes relatively to the first. Let the planes $X'_1\, OX'_2$ and $X_1\, OX_2$ intersect in ON (Fig. 1). We make a definite choice of direction along this line and let α be the angle $X'_1\, ON$, reckoned from OX'_1. We also introduce the angles $\beta = X'_3\, OX_3$ and $\gamma = NOX_1$. These three angles completely characterize the position of the second system relative to the first, i.e. they completely characterize the rotation which we shall denote by the symbol $\{\alpha, \beta, \gamma\}$. It follows at once from the above that our motion is the result of consecutively carrying out the following three motions: (1) rotation by angle α about the axis X'_3; (2) rotation by the angle β about the new position of X'_1; (3) rotation by angle γ about the new axis X_3. These three angles are generally known as Euler's angles, and we can write their limits of variation as follows:

$$0 \leqslant \alpha < 2\pi; \quad 0 \leqslant \beta < \pi; \quad 0 \leqslant \gamma < 2\pi.$$

If $\beta = 0$, the motion $\{\alpha, \beta, \gamma\}$ simply reduces to a rotation by the angle $\alpha + \gamma$ about axis X_3, and we have in this sense for any δ:

$$\{\alpha, 0, \gamma\} = \{\alpha + \delta, 0, \gamma - \delta\}.$$

This shows that, for certain cases of parameters $\{\alpha, \beta, \gamma\}$, the rotation of the space about the origin is not single-valued, in other words, the same rotation corresponds to different values of the parameters. Expressions may readily be deduced for the coefficients a_{ik} in terms of trigonometric functions of the angles α, β, γ [cf. 62]. We shall deal later with a further choice for the parameters characterizing a rotation of space about the origin, and we shall also return to Euler's angles.

21. General linear transformations of real three-dimensional space. We shall now consider real linear transformations of type (2) with arbitrary coefficients, though it will always be assumed that the transformation determinant differs from zero:

$$|a_{ik}| \neq 0. \tag{7}$$

The transformation is usually said to be *non-singular* in this case. If it does not satisfy conditions (5), it is related to a deformation of space [II, 113]. It should be noticed that the characteristic feature of (2) is the matrix of coefficients which implies a fully defined rule for passing from any vector with components (x_1, x_2, x_3) to a new vector with components (x_1', x_2', x_3'). We shall use a single letter to denote the total matrix:

$$A = \begin{Vmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31}, & a_{32}, & a_{33} \end{Vmatrix}, \tag{8}$$

the matrix being written between double strokes as before to distinguish it from the determinant. We shall write the determinant of matrix (8) as $D(A)$. This is some determinate number. We shall write transformation (2) symbolically as

$$\mathbf{x}' = A\mathbf{x}, \tag{9}$$

where $\mathbf{x}'$ is the vector with components (x_1', x_2', x_3') and $\mathbf{x}$ has components (x_1, x_2, x_3).

The *identity transformation* is that in which every vector remains unchanged; the matrix corresponding to this is

$$\begin{Vmatrix} 1, & 0, & 0, \\ 0, & 1, & 0, \\ 0, & 0, & 1, \end{Vmatrix} \tag{10}$$

which is generally known as the *unit matrix* and is denoted by the symbol I.

Assuming $D(A) \neq 0$, we can solve equations (2) with respect to (x_1, x_2, x_3) and arrive at the expressions

$$\left.\begin{aligned} x_1 &= \frac{A_{11}}{D(A)}x_1' + \frac{A_{21}}{D(A)}x_2' + \frac{A_{31}}{D(A)}x_3' \\ x_2 &= \frac{A_{12}}{D(A)}x_1' + \frac{A_{22}}{D(A)}x_2' + \frac{A_{32}}{D(A)}x_3' \\ x_3 &= \frac{A_{13}}{D(A)}x_1' + \frac{A_{23}}{D(A)}x_2' + \frac{A_{33}}{D(A)}x_3' \end{aligned}\right\}, \tag{11}$$

where the A_{ik} are the cofactors of the a_{ik} in $D(A)$. This linear transformation is usually referred to as the inverse of (2), and if A denotes the matrix of (2), the matrix of (11) is written A^{-1}. We now introduce a concept of importance for what follows, that of the product of two transformations or of two matrices. Let us have two linear transformations, from (x_1, x_2, x_3) to (x_1', x_2', x_3'):

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ x_2' &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x_3' &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned}\right\} \quad \text{or} \quad \mathbf{x}' = A\mathbf{x} \tag{12}$$

then from (x_1', x_2', x_3') to (x_1'', x_2'', x_3''):

$$\left.\begin{aligned} x_1'' &= b_{11}x_1' + b_{12}x_2' + b_{13}x_3' \\ x_2'' &= b_{21}x_1' + b_{22}x_2' + b_{23}x_3' \\ x_3'' &= b_{31}x_1' + b_{32}x_2' + b_{33}x_3' \end{aligned}\right\} \quad \text{or} \quad \mathbf{x}'' = B\mathbf{x}'. \tag{13}$$

These successive transitions from (x_1, x_2, x_3) to (x_1', x_2', x_3') then from (x_1', x_2', x_3') to (x_1'', x_2'', x_3'') can be replaced by a direct transition from (x_1, x_2, x_3) to (x_1'', x_2'', x_3''), this latter being also a linear transformation:

$$x_k'' = c_{k1}x_2 + c_{k2}x_2 + c_{k3}x_3 \quad (k = 1, 2, 3). \tag{14}$$

This last transformation is described as the product of transformations (12) and (13), an essential point to notice being the order in which the transformations are carried out. We obtain (14) by substituting from (12) in the right-hand sides of (13). This gives us expressions for the elements c_{ik} of the transformation product in terms of the elements of the original transformations:

$$c_{ik} = \sum_{s=1}^{3} b_{is}a_{sk} \qquad (i, k = 1, 2, 3). \tag{15}$$

We usually write (14) as follows:

$$\mathbf{x}'' = BA\mathbf{x}. \tag{16}$$

The matrix C with elements c_{ik} as given by (15) is called the product of matrices A and B and is written thus:

$$C = BA, \tag{17}$$

where the product must be read from right to left in the sense of the order of carrying out the transformations. If we make use of (15) and the multiplication theorem for determinants, we can write the obvious equality:

$$D(C) = D(B)\, D(A), \tag{18}$$

i.e. *the determinant of the product of two transformations is equal to the product of their determinants.* We can easily prove the following relationship, which has a simple geometrical meaning:

$$AA^{-1} = A^{-1}A = I. \tag{19}$$

We also notice that it follows from the actual derivation of the inverse transformation that the inverse of A^{-1} is the transformation A. For, on solving system (11) with respect to the x_k', we obviously again get expressions (2). We can write this as follows:

$$(A^{-1})^{-1} = A. \tag{20}$$

The concept of transformation product can be extended to the case of any number of factors; e.g. the result of successive transformations with matrices A, B, and C is a transformation with matrix D:

$$D = CBA. \tag{21}$$

If matrices A, B, C have the elements

$$a_{ik},\ b_{ik} \text{ and } c_{ik},$$

the matrix D will have elements given by the expressions:

$$d_{ik} = \sum_{p,\,q=1}^{3} c_{iq} b_{qp} a_{pk}. \tag{22}$$

We have, in fact, for the elements of the matrix $E = BA$:

$$e_{ik} = \sum_{p=1}^{3} b_{ip} a_{pk}$$

and finally, for the element of CE, by (15):

$$d_{ik} = \sum_{q=1}^{3} c_{iq} e_{ak},$$

whence (22) follows. It must be mentioned that the elements of a matrix A will often be written in future as

$$\{A\}_{ik}.$$

Matrix products are not generally subject to the *commutative law* i.e. they change when factors are interchanged, so that in general e.g. $BA \neq AB$. They obey the associative law, however, that is to say, their factors can be grouped:

$$C(BA) = (CB)A. \tag{23}$$

On the left-hand side, we must multiply A by B, then multiply the result by C. On the right-hand side, we first multiply B by C, then multiply A by the result. It is easily seen that in both cases the elements of the matrix finally obtained are given by (22). This has already been proved for the left-hand side; we have for the right-hand side, on carrying out the successive multiplications:

$$\{CB\}_{ik} = \sum_{q=1}^{3} \{C\}_{iq}\{B\}_{qk}$$

and

$$\{(CB)A\}_{ik} = \sum_{p=1}^{3} \{CB\}_{ip}\{A\}_{pk} = \sum_{p,q=1}^{3} \{C\}_{iq}\{B\}_{qp}\{A\}_{pk}$$

which is evidently (22) in our new notation.

A further important type of linear transformation must be mentioned, in which we have:

$$x'_1 = k_1 x_1; \quad x'_2 = k_2 x_2; \quad x'_3 = k_3 x_3, \tag{24}$$

and which amounts to extension (or contraction) along the coordinate axes, the extension being characterized by the numerical coefficients k_1, k_2, k_3. The matrix of this transformation is obviously

$$\begin{Vmatrix} k_1, & 0, & 0 \\ 0, & k_2, & 0 \\ 0, & 0, & k_3 \end{Vmatrix},$$

i.e. all the elements not on the principal diagonal are zero. We refer to this type as a *diagonal matrix*, and denote it by

$$[k_1 \ k_2, \ k_3].$$

In particular, if $k_1 = k_2 = k_3$, the transformation reduces to multiplication of all the components of a vector by the same number k and is

evidently the transformation of similitude with centre at the origin. Every vector changes its length, which is multiplied by k, without changing its direction (we are assuming $k > 0$). The following simple notation is used in this case:

$$\mathbf{x}' = k\mathbf{x},$$

i.e. we look on the number k as a particular case of a matrix, and take it as in fact the diagonal matrix with the same element k on the principal diagonal:

$$\begin{Vmatrix} k, & 0, & 0 \\ 0, & k, & 0 \\ 0, & 0, & k \end{Vmatrix}. \tag{25}$$

It may readily be seen by using (15) that multiplication of such matrices reduces to the ordinary cross-multiplication of numbers:

$$[k, k, k] \cdot [l, l, l] = [kl, kl, kl].$$

It may easily be shown that, in general, the simple multiplication rule:

$$[k_1, k_2, k_3] \cdot [l_1, l_2, l_3] = [k_1 l_1, k_2 l_2, k_3 l_3], \tag{26}$$

applies for diagonal matrices, i.e. two extensions along the coordinate axes are equivalent to a single extension with coefficients equal to the products of the corresponding coefficients of the component extensions. An immediate consequence of (26) is that the product of two diagonal matrices is unchanged on interchanging the factors. By using (15) and representation (25) of a number as a diagonal matrix, it can easily be seen that the product kA is obtained simply by multiplying all the elements of A by the number k. This product is independent of the order of the factors, i.e.

$$\{ kA \}_{ik} = \{ Ak \}_{ik} = k \{ A \}_{ik}. \tag{27}$$

We have regarded the basic linear transformation (2) as a deformation of space in which a vector with components (x_1, x_2, x_3) becomes a new vector with components (x_1', x_2', x_3'). Of course (2) can also be interpreted, as already mentioned, as a point transformation in which a point with coordinates (x_1, x_2, x_3) becomes a point with coordinates (x_1', x_2', x_3').

We could have used any system of axes, in other words, any fundamental vector set, for defining vector components, i.e. we could have taken any three non-coplanar unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ as the fundamental

set, in which case, as we know from [II, 102], any vector $\mathbf{x}$ can be expressed uniquely in the form

$$\mathbf{x} = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k}. \tag{28}$$

The numbers x_1, x_2, x_3 are called the components of $\mathbf{x}$ in the coordinate system defined by the fundamental set $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Our next task is to see the effect of a different choice of fundamental set on the form of the linear transformation.

More precisely, if a linear transformation in the coordinate system defined by $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is given by (12), what is the form of this same transformation of space in another coordinate system, defined by say $\mathbf{i}_1, \mathbf{j}_1, \mathbf{k}_1$? Let the new fundamental set be given in terms of the old by the expressions:

$$\left.\begin{aligned} \mathbf{i}_1 &= t_{11} \mathbf{i} + t_{12} \mathbf{j} + t_{13} \mathbf{k} \\ \mathbf{j}_1 &= t_{21} \mathbf{i} + t_{22} \mathbf{j} + t_{23} \mathbf{k} \\ \mathbf{k}_1 &= t_{31} \mathbf{i} + t_{32} \mathbf{j} + t_{33} \mathbf{k}. \end{aligned}\right\} \tag{29}$$

It will be noticed that the determinant made up of coefficients t_k^i cannot vanish; if it did, $\mathbf{i}_1, \mathbf{j}_1, \mathbf{k}_1$ would be linearly dependent, i.e. coplanar. In the new coordinate system, the vector given by (28) will have new components:

$$y_1 \mathbf{i}_1 + y_2 \mathbf{j}_1 + y_3 \mathbf{k}_1.$$

We first of all establish expressions for the new components in terms of the old. We obviously have, on substituting expressions (29) for the new fundamental vectors:

$$\sum_{s=1}^{3} y_s (t_{s1} \mathbf{i} + t_{s2} \mathbf{j} + t_{s3} \mathbf{k}) = x_1 \mathbf{i} + x_2 \mathbf{j} + x_3 \mathbf{k}.$$

We get expressions for the old components in terms of the new by equating coefficients of $\mathbf{i}, \mathbf{j}, \mathbf{k}$:

$$\left.\begin{aligned} x_1 &= t_{11} y_1 + t_{21} y_2 + t_{31} y_3 \\ x_2 &= t_{12} y_1 + t_{22} y_2 + t_{32} y_3 \\ x_3 &= t_{13} y_1 + t_{23} y_2 + t_{33} y_3. \end{aligned}\right\} \tag{30}$$

The first subscript remains unchanged along a row in the matrix of transformation (29), whereas the second subscript remains unchanged in the rows of the matrix of (30). The matrices thus differ to the extent of rows being replaced by columns. If T denotes the matrix of (29), the array of (30) is known as the transpose of T and is written T^*.

Expressions (30) may be written in the abbreviated form:

$$(x_1, x_2, x_3) = T^{(*)}(y_1, y_2, y_3), \tag{31}$$

where (x_1, x_2, x_3) are the three components of a vector with respect to the first fundamental set, and (y_1, y_2, y_3) the components with respect to the new fundamental set. Conversely, the new components may be expressed in terms of the old by

$$(y_1, y_2, y_3,) = T^{(*)-1}(x_1, x_2, x_3).$$

where T^{*-1} is the inverse linear transformation to T^*. This is generally known as the *contragredient* of T. For brevity, we use a special letter to denote the array corresponding to it:

$$U = T^{(*)-1} \tag{32}$$

We can thus say that a change in the fundamental set in accordance with (29) implies that the components of every vector undergo a linear transformation with the matrix U defined by (32). Hence the two vectors $\mathbf{x}(x_1, x_2, x_3)$ and $\mathbf{x}'(x'_1, x'_2, x'_3)$ that appear in transformation (9) will have different components after transformation of the fundamental set, these being given in terms of the original components by

$$(y_1, y_2. y_3) = U(x_1. x_2, x_3);\ (y'_1, y'_2, y'_3) = U(x'_1, x'_2, x'_3). \tag{33}$$

Our problem is to establish the linear relationship between components (y_1, y_2, y_3) and (y'_1, y'_2, y'_3). We can pass from the former to the latter by the following method: we first pass from vector (y_1, y_2, y_3) to (x_1, x_2, x_3) with the aid of the matrix U^{-1}, by (33). Then we pass from (x_1, x_2, x_3) to (x', x', x') with the aid of matrix A of (9), and finally, from (x'_1, x'_2, x'_3) to (y'_1, y'_2, y'_3) with the aid of matrix U. We thus end up with the linear transformation:

$$\mathbf{y}' = UAU^{-1}\mathbf{y}. \tag{34}$$

This transformation is said to be similar to transformation (9), and its matrix UAU^{-1} is said to be similar to matrix A.

Our result may finally be stated as: *if the linear transformations of vector components due to a change in the fundamental set are given by expressions* (33), *any linear spatial transformation with the form in the original fundamental set:*

$$\mathbf{x}' = A\mathbf{x},$$

becomes in the new coordinate system:

$$\mathbf{y}' = UAU^{-1}\mathbf{y}.$$

22. Covariant and contravariant affine vectors. Suppose that linear transformation (9) simply expresses the passage from one system of Cartesian axes to another, i.e. its coefficients are the direction-cosines given in table (3). In this case, as we saw in [20], the transpose $A^{(*)}$ is the same as the inverse A^{-1}, and the contragradient $A^{(*)-1}$ is therefore the same as the original matrix A, i.e.

$$A^{(*)} = A^{-1}; \quad A^{(*)-1} = A. \tag{35}$$

If we take a vector of constant length and direction, we can naturally say that its components are transformed in accordance with the same expressions (9) as the coordinates, i.e.

$$\left.\begin{aligned} x_1' &= a_{11} x_1 + a_{12} x_2 + a_{13} x_3 \\ x_2' &= a_{21} x_1 + a_{22} x_2 + a_{23} x_3 \\ x_3' &= a_{31} x_1 + a_{32} x_2 + a_{33} x_3. \end{aligned}\right\} \tag{36}$$

We can, therefore, say that a vector is completely characterized by three numbers in any fixed Cartesian system, and on passage from one Cartesian system to another the three numbers (vector components) are transformed in accordance with the same expressions (36) as the coordinates. Suppose that we now take into account not only passage from one Cartesian system to another, but all the generally possible linear transformations of coordinates with non-zero determinants which corresponds, as we saw above, to an arbitrary choice of three non-coplanar vectors as the fundamental set. As above, along with matrix A of transformation (36), we shall consider the contragradient $V = A^{(*)-1}$. These are distinct in the general case, so that we have two possibilities for defining a vector in any linear coordinate transformation. In the first place, we can define a vector as a set of three numbers which is transformed on passage from one coordinate system to another by the same formulae as the coordinates themselves, i.e. by

$$(x_1', x_2', x_3') = A\,(x_1, x_2, x_3). \tag{37}$$

Such a vector is described as a *contravariant affine vector*, the general linear transformation (36) being sometimes referred to as an affine transformation. Alternatively, we can define a vector such that its components undergo the corresponding contragradient transformation for any linear transformation (36), i.e.

$$(x_1', x_2', x_3') = V\,(x_1, x_2, x_3). \tag{38}$$

Such a vector is known as a *covariant affine vector*.

In both cases, given the components of a vector in any one coordinate system, we automatically obtain the components in any other coordinate system which is derived from the original system by means of an affine transformation. Examples of both types of vector are as follows. The radius vector joining two given points in space is clearly contravariant, since its components in the above sense of the word (the differences between the coordinates of its end-points) are transformed in accordance with the same linear formulae as the coordinates themselves. Another example of a contravariant vector may

be quoted. We take the coordinates (x_1, x_2, x_3) of a point as functions of a parameter t and define a velocity vector with components

$$\left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \frac{dx_3}{dt}\right).$$

On differentiating the basic expressions (36) with respect to t, we see at once that the velocity vector is contravariant.

We now give an example of a covariant vector. Let $f(x_1, x_2, x_3)$ be a function of a point in space; the *gradient* of the function in any given coordinate system is defined as the vector with components

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}.$$

We have by (36) and the rule for differentiating functions of a function:

$$\frac{\partial f}{\partial x_s} = a_{1s}\frac{\partial f}{\partial x'_1} + a_{2s}\frac{\partial f}{\partial x'_2} + a_{3s}\frac{\partial f}{\partial x'_3} \quad (s = 1, 2, 3),$$

i.e. the components of the gradient along the (x_1, x_2, x_3) axes are given in terms of the components along the (x'_1, x'_2, x'_3) axes by a linear transformation with matrix $A^{(*)}$, whence it follows that the components along the (x'_1, x'_2, x'_3) axes are given in terms of the components along the (x_1, x_2, x_3) axes by a linear transformation with matrix $A^{(*)-1} = V$, i.e. the gradient of a function is in fact a covariant vector.

Expressions (37) and (38) may readily be written in terms of the partial derivatives of the new coordinates with respect to the old, and vice versa. We first introduce a notation which is somewhat different to the above and is more usual in vector theory: a superscript is used for the components of contravariant vectors and a subscript for the covariant components, the corresponding coordinates being themselves denoted by a superscript.

The coefficients of transformation (36) can be written as follows in terms of the partial derivatives:

$$a_{ik} = \frac{\partial x'^{(i)}}{\partial x^{(k)}}. \tag{39}$$

The elements of the contragradient matrix V become:

$$V_{ik} = \frac{A_{ik}}{D(A)},$$

and $(A^{-1})^{(*)}$ has the same elements, i.e.

$$A^{(*)-1} = (A^{-1})^{(*)},$$

i.e. we can first pass to the inverse matrix then interchange rows and columns. On passage to the inverse matrix, the coefficient c_{ik} becomes $\partial x^{(i)}/\partial x'^{(k)}$, and after transposition, we have for the elements of matrix V:

$$V_{ik} = \frac{\partial x^{(k)}}{\partial x'^{(i)}}. \tag{40}$$

Let the components of a contravariant vector be $u^{(s)}$ in coordinates $x^{(k)}$

and $u'^{(s)}$ in coordinates $x'^{(s)}$. We have by definition:

$$u'^{(i)} = \sum_{s=1}^{3} \frac{\partial x'^{(i)}}{\partial x^{(s)}} u^{(s)} \qquad (i = 1, 2, 3). \tag{41}$$

Similarly, we get by definition for a covariant vector:

$$u'_i = \sum_{s=1}^{3} \frac{\partial x^{(s)}}{\partial x'^{(i)}} u_s. \tag{42}$$

It may be mentioned that these formulae can be used for defining the components of a vector not only on linear transformation of the coordinates but with the most general type of transformation, when the individual coordinates are expressed in terms of others with the aid of in general non-linear functions.

We shall indicate another possible definition of covariant vector, when the contravariant vector is defined simply as the vector whose components are transformed in accordance with the same formulae as the coordinates. Let $u^{(s)}$ be a given contravariant, and v_s a covariant vector.

We form the sum:

$$u^{(1)} v_1 + u^{(2)} v_2 + u^{(3)} v_3. \tag{43}$$

This may easily be seen to remain invariable, or in other words, to be a scalar, if $u^{(s)}$ and v_s vary in accordance with the corresponding expressions (41) and (42).

For the rule for differentiating functions of a function gives us at once:

$$\sum_{s=1}^{3} u'^{(s)} v'_s = \sum_{s=1}^{3} \left[\sum_{k=1}^{3} \frac{\partial x'^{(s)}}{\partial x^{(k)}} u^{(k)} \right] \left[\sum_{l=1}^{3} \frac{\partial x^{(l)}}{\partial x'^{(s)}} v_l \right] = u^{(1)} v_1 + u^{(2)} v_2 + u^{(3)} v_3.$$

Hence, having defined a contravariant vector by the method given above, we can find the transformation rule for the covariant components from the requirement that sum (43) remains invariable. An exact repetition of the working of the previous section leads us to the conclusion that, given the invariability of (43), the components v_s must undergo linear transformation contragradient to that suffered by the components $u^{(s)}$. We suggest that the reader show that, for any (linear or non-linear) coordinate transformation, the velocity vector is a contravariant whilst the gradient of a function is a covariant vector.

A distinction is worth noticing between contravariant and covariant vectors which have been defined in a purely formal manner above, by formulae for passing from one system to the other. Let $\mathbf{x}$ be a vector of given length and direction. Given the fundamental set, we form components in accordance with (28) and now refer to them as the *contravariant components*, (28) being written in the form

$$\mathbf{x} = x^{(1)} \mathbf{i} + x^{(2)} \mathbf{j} + x^{(3)} \mathbf{k}. \tag{44}$$

The *covariant component* of $\mathbf{x}$ along $\mathbf{i}$ is defined as the rectangular projection of $\mathbf{x}$ on $\mathbf{i}$, multiplied by the length of $\mathbf{i}$, and similarly for the other fundamental vectors. We thus have, for each fundamental set, three covariant components (x_1, x_2, x_3). It can be shown that these are transformed like the components

of a covariant vector on passage from one fundamental set to another. For it can be shown (we shall not dwell on the proof) that the expression

$$x^{(1)} x_1 + x^{(2)} x_2 + x^{(3)} x_3$$

here gives the square of the length of vector $\mathbf{x}$ and is therefore unchanged on transformation of the fundamental set.

23. Tensors. We now turn to a generalization of vectors, only linear coordinate transformations being considered initially. Let the array of nine numbers

$$b_{ik} \quad (i, k = 1, 2, 3)$$

be given in a certain coordinate system.

We form the expression

$$\sum_{i,k=1}^{3} b_{ik} u^{(i)} v^{(k)}, \tag{45}$$

where $u^{(i)}$ and $v^{(k)}$ are the components of two contravariant vectors. On passing to new coordinates, we can express the $u^{(i)}$ and $v^{(k)}$ in (45) in terms of the new components $u'^{(i)}$ and $v'^{(k)}$ and hence transform (45) as follows:

$$\sum_{i,k=1}^{3} b_{ik} u^{(i)} v^{(k)} = \sum_{i,k=1}^{3} b'_{ik} u'^{(i)} v'^{(k)}. \tag{46}$$

Now we still have an array of nine numbers, with elements b'_{ik}, in the new coordinate system. Such an array, defined in any coordinate system by requiring invariance of expression (45), is described as a *covariant tensor of the second rank*. Similarly, on taking two covariant vectors with components u_i and v_k and forming the expression

$$\sum_{i,k=1}^{3} b^{(i,k)} u_i v_k, \tag{47}$$

where the array of nine numbers $b^{(i,k)}$ is specified in some given coordinate system, we obtain a similar array in any other coordinate system on requiring invariance of expression (47). We now have a *contravariant tensor of the second rank*. Finally, if we take a contravariant vector with components $u^{(i)}$ and a covariant vector with components v_k and form the expression

$$\sum_{i,k=1}^{3} b_i^{(k)} u^{(i)} v_k, \tag{48}$$

we arrive in precisely the same way at a *mixed tensor of the second rank*.

We now show how, given the coefficients of the linear coordinate transformation (36), expressions can be derived for the components of a tensor in the new coordinates in terms of its components in the old. We start by considering a covariant tensor of the second rank. The components $u^{(i)}$ and $v^{(k)}$ of the contravariant vectors in the old coordinates are expressed in terms of the components $u'^{(i)}$ and $v'^{(k)}$ in the new coordinate system with the aid of a linear trans-

formation of matrix A^{-1}. This gives us, on writing the elements of the matrix as $\{A^{-1}\}_{ik}$:

$$u^{(i)} = \sum_{k=1}^{3} \{A^{-1}\}_{ik} u'^{(k)}; \quad v^{(i)} = \sum_{k=1}^{3} \{A^{-1}\}_{ik} v'^{(k)}.$$

We substitute in (45), find the coefficients of the products $u'^{(i)} v'^{(k)}$, and thus obtain for the components b'_{ik} of the tensor in the new coordinate system:

$$b'_{ik} = \sum_{p, \ =1}^{3} b_{pq} \{A^{-1}\}_{pi} \{A^{-1}\}_{qk}. \tag{49}$$

Similarly, for a contravariant tensor of the second rank, we have to express the components of the covariant vectors u_i and v_k in terms of the new components. By definition of contravariant vector, u'_i is given in terms of u_i by means of the array $A^{(*)-1}$, so that u_i is given in terms of u via the array $A^{(*)}$, the transpose of A, and similarly for v_i:

$$u_i = \sum_{k=1}^{3} \{A\}_{ki} u'_k; \quad v_i = \sum_{k=1}^{3} \{A\}_{ki} v'_k.$$

Substitution in (47) gives us the transformation for the components of a contravariant tensor of the second rank:

$$b'^{(i, k)} = \sum_{p, q=1}^{3} b^{(p, q)} \{A\}_{ip} \{A\}_{kq}. \tag{50}$$

Similarly, we have the following transformation for the components of a mixed tensor of the second rank:

$$b_i'^{(k)} = \sum_{p, q=1}^{3} b_p^{(q)} \{A^{-1}\}_{pi} \{A\}_{kq}. \tag{51}$$

If we express the coefficients of the linear transformation in terms of the partial derivatives

$$\frac{\partial x'^{(i)}}{\partial x^{(k)}} \text{ and } \frac{\partial x^{(i)}}{\partial x'^{(k)}}$$

and substitute these expressions in the above formulae, we get formulae for transforming tensors of the second rank in the case of any coordinate transformation. Analogous definitions to the above are possible for tensors of rank higher than the second, but we shall not dwell on this.

We have constantly been concerned above with matrices expressing linear transformations of three-dimensional space with given coordinate axes. Let such a matrix be B and let an affine coordinate transformation have been carried out in accordance with

$$(y_1, y_2, y_3) = A\,(x_1, x_2, x_3),$$

where A is a matrix with a non-zero determinant. As we have shown above, our spatial transformation has the matrix in the new coordinates:

$$ABA^{-1}.$$

It is easily seen that the transformation worked out above for a mixed tensor of the second rank has the same matrix. For, on applying the rule for matrix multiplication, we have

$$\{BA^{-1}\}_{ql} = \sum_{p=1}^{3} \{B\}_{qp} \{A^{-1}\}_{pl},$$

followed by:

$$\{A\,(BA^{-1})\}_{ki} = \sum_{q=1}^{3} \{A\}_{kq} \{BA^{-1}\}_{qi} = \sum_{p,\,q=1}^{3} \{B\}_{qp} \{A^{-1}\}_{pi} \{A\}_{kq}.$$

If we write $b_k^{(i)}$ instead of $\{B\}_{ik}$, we have an expression of the same form as (51). The matrix of a linear transformation of space is thus a mixed tensor of the second rank.

Some tensors of a particular kind must be mentioned. Let a covariant tensor in a given coordinate system have the property that

$$b_{ik} = b_{ki} \quad (i, k = 1, 2, 3). \tag{52}$$

It may easily be seen to have the same property in any other coordinate system. For, by (49):

$$b'_{ki} = \sum_{p,\,q=1}^{3} b_{pq} \{A^{-1}\}_{pk} \{A^{-1}\}_{qi}$$

or by (52):

$$b'_{ki} = \sum_{p,\,q=1}^{3} b_{qp} \{A^{-1}\}_{pk} \{A^{-1}\}_{qi},$$

or, on changing the notation for the variables of summation:

$$b'_{ki} = \sum_{p,\,q=1}^{3} b_{pq} \{A^{-1}\}_{qk} \{A^{-1}\}_{pi},$$

whence it is clear that b'_{ki} is in fact the same as b'_{ik}. We describe this type as a *symmetrical covariant tensor*. A *symmetrical contravariant tensor* can be defined in exactly the same way. Similarly, if $b_{ik} = -b_{ik}$ or $b^{(i,k)} = -b^{(k,i)}$ in one coordinate system, the same will be true in any coordinate system, and the corresponding tensor is said to be *skew-symmetric*. The same situation does not hold for a mixed tensor, so that, e.g., $b_i^{(k)} = b_k^{(i)}$ is not an invariant relationship on transformation of the coordinates. We shall consider next some particular cases of tensors.

24. Examples of affine orthogonal tensors. We confine ourselves in future examples to the linear coordinate transformations discussed in [20], which correspond to the passage from one Cartesian system to another, and which are generally known as orthogonal transformations of three-dimensional space. The contragradient transformation $A^{(*)-1}$ coincides with A for these, as we have already seen, and the distinction between contravariant and covariant vectors disappears. Similarly, it is clear that we now have just the one concept of second rank tensor. If we write $\{A\}_{ik}$ for the matrix coefficients of an ortho-

gonal coordinate transformation as above, we have the following formula for transforming a tensor of the second rank:

$$b'_{ik} = \sum_{p,\, q=1}^{3} b_{pq} \{A\}_{ip} \{A\}_{kq}\,, \tag{53}$$

which follows at once from the expressions of the previous section. The elements of a column of $\| b_{ik} \|$ will be looked on as vector components. Hence we have the three vectors:

$$\mathbf{b}^{(1)}\,(b_{11}, b_{21}, b_{31});\quad \mathbf{b}^{(2)}\,(b_{12}, b_{22}, b_{32});\quad \mathbf{b}^{(3)}\,(b_{13}, b_{23}, b_{33}).$$

We shall say that the first of these corresponds to the x_1, the second to the x_2, and the third to the x_3 axis. We now relate a vector $\mathbf{b}^{(n)}$ to any direction n in accordance with the formula:

$$\mathbf{b}^{(n)} = \cos(n, x_1)\,\mathbf{b}^{(1)} + \cos(n, x_2)\,\mathbf{b}^{(2)} + \cos(n, x_3)\,\mathbf{b}^{(3)}. \tag{54}$$

We next replace the original Cartesian system (x_1, x_2, x_3) by (x'_1, x'_2, x'_3) and use (54) to form the vectors corresponding to the new directions of the axes:

$$\mathbf{b}'^{(k)} = \cos(x'_k, x_1)\,\mathbf{b}^{(1)} + \cos(x'_k, x_2)\,\mathbf{b}^{(2)} + \cos(x'_k, x_3)\,\mathbf{b}^{(3)}. \tag{55}$$

On taking the projections of these vectors on the new (x'_1, x'_2, x'_3) axes, we get an array of nine numbers $\| b'_{ik} \|$ analogous to $\| b_{ik} \|$. We show that the elements of the new array are given in terms of those of the original array precisely by the formulae for transforming tensors of the second rank. For, taking say the element b'_{12}, this is by definition the component of the vector $\mathbf{b}'^{(2)}$ along the new x'_1 axis, and (55) gives

$$\mathbf{b}'^{(2)} = \cos(x'_2, x_1)\,\mathbf{b}^{(1)} + \cos(x'_2, x_2)\,\mathbf{b}^{(2)} + \cos(x'_2, x_3)\,\mathbf{b}^{(3)}, \tag{56}$$

so that $\mathbf{b}'^{(2)}$ is clearly a linear function of the vectors $\mathbf{b}^{(i)}$. All we have to do to get b'_{12} is to replace the $\mathbf{b}^{(i)}$ on the right-hand side of (56) by their projections on the x'_1 axis, i.e. by the following expressions:

$$\mathbf{b}^{(i)} \text{ — by } b_{1i}\cos(x'_1, x_1) + b_{2i}\cos(x'_1, x_2) + b_{3i}\cos(x'_1, x_3) \quad (i = 1, 2, 3).$$

We now notice that, in accordance with table (3):

$$\cos(x_i, x_k) = a_{ik} = \{A\}_{ik}.$$

We have on making these substitutions for the vectors on the right-hand side of (56):

$$b'_{12} = \sum_{p,\, q=1}^{3} b_{pq} \{A\}_{1p} \{A\}_{2q}$$

which is precisely the same as (53). We can therefore assert that, if three vectors $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \mathbf{b}^{(3)}$ are defined for three mutually perpendicular directions and a vector for any direction (n) is defined by (54), the array of the nine numbers giving the projections of the vectors $\mathbf{b}'^{(k)}$ $(k = 1, 2, 3)$ on the $x'^{(k)}$ axes in any Cartesian system defines an affine orthogonal tensor of the second rank, i.e. a tensor of the second rank, defined for all possible orthogonal transformations.

It must be noted that, when we speak of $\mathbf{b}^{(1)}$ corresponding to the direction of a given x_1 axis, this does not mean that $\mathbf{b}^{(1)}$ must be directed along the x_1 axis. What is essential is expression (54), which relates the vector $\mathbf{b}^{(n)}$ to any direction (n), where the direction of the vector does not in general coincide with (n).

Two examples of affine orthogonal tensors of the second rank may be quoted. The first of these is the stress tensor, familiar in the theory of elasticity. We consider an infinitesimal surface $d\sigma$ with normal (n) at a fixed point M of an elastic body under deformation. The action on the surface of the part of the elastic medium lying on the side defined by the normal direction is taken in the theory of elasticity to be equivalent to the product of a vector $\mathbf{b}^{(n)}$, dependent on the direction (n), with the magnitude $d\sigma$ of the area. By considering the equilibrium conditions for an infinitesimal tetrahedron cut from the body, we arrive at equation (54), which shows at once that the stress is a tensor of the second rank. This tensor is given in any Cartesian system by an array of nine numbers $\| b_{ik} \|$, the tensor being symmetrical, i.e. $b_{ik} = b_{ki}$, as shown in the theory of elasticity. In other words, the projection on the x_i axis of the stress acting on an area perpendicular to the x_k axis is equal to the projection on the x_k axis of the stress acting on an area perpendicular to the x_i axis.

We now turn to the second example. Let $\mathbf{C}(M)$ be a vector field. If we choose a Cartesian system (x_1, x_2, x_3) and take the derivatives of the field components (c_1, c_2, c_3) with respect to the coordinates, we get the following array of nine quantities:

$$\left\| \begin{matrix} \dfrac{\partial c_1}{\partial x_1}, & \dfrac{\partial c_1}{\partial x_2}, & \dfrac{\partial c_1}{\partial x_3} \\ \dfrac{\partial c_2}{\partial x_1}, & \dfrac{\partial c_2}{\partial x_2}, & \dfrac{\partial c_2}{\partial x_3} \\ \dfrac{\partial c_3}{\partial x_1}, & \dfrac{\partial c_3}{\partial x_2}, & \dfrac{\partial c_3}{\partial x_3} \end{matrix} \right\|, \tag{57}$$

We define a corresponding vector $\partial \mathbf{c}/\partial n$ for any given direction (n); for instance, the elements of the kth column of (57) give the components of the vector that corresponds to the x_k axis. We have the expression, for any direction (n) [II, 108]:

$$\frac{\partial c_i}{\partial n} = \cos(n, x_1)\frac{\partial c_i}{\partial x_1} + \cos(n, x_2)\frac{\partial c_i}{\partial x_2} + \cos(n, x_3)\frac{\partial c_i}{\partial x_3} \quad (i = 1, 2, 3),$$

i.e. array (57) defines a tensor of the second rank. This tensor is in general neither symmetric nor anti-symmetric. But it is readily expressible as the sum of a symmetric and anti-symmetric tensor, where the sum of two matrices is understood to mean the matrix consisting of the sums of corresponding elements.

We shall first make a general preliminary remark. The linearity of expression (53) implies that, if $\| b_{ik} \|$ and $\| c_{ik} \|$ are two tensors, the sum $\| b^{ik} + c_{ik} \|$ is likewise a tensor. Furthermore, the same formula remains valid on interchange of the subscripts, i.e.

$$b'_{ki} = \sum_{p, q=1}^{3} b_{qp} \{A\}_{ip} \{A\}_{kq},$$

so that if a matrix defined for any axes yields a tensor, its transpose likewise yields a tensor. Suppose now that we are given the tensor $\| b_{ik} \|$.

We can write this as a sum:

$$\| b_{ik} \| = \left\| \frac{b_{ik} + b_{ki}}{2} \right\| + \left\| \frac{b_{ik} - b_{ki}}{2} \right\|.$$

The first term on the right is clearly a symmetric tensor, and the second term an anti-symmetric tensor.

On applying this decomposition to the tensor defined by (57), we get for its symmetric part:

$$\left\| \begin{matrix} \dfrac{\partial c_1}{\partial x_1}, & \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_2} + \dfrac{\partial c_2}{\partial x_1}\right), & \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_3} + \dfrac{\partial c_3}{\partial x_1}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_2} + \dfrac{\partial c_2}{\partial x_1}\right), & \dfrac{\partial c_2}{\partial x_2}, & \dfrac{1}{2}\left(\dfrac{\partial c_2}{\partial x_3} + \dfrac{\partial c_3}{\partial x_2}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_3} + \dfrac{\partial c_3}{\partial x_1}\right), & \dfrac{1}{2}\left(\dfrac{\partial c_2}{\partial x_3} + \dfrac{\partial c_3}{\partial x_2}\right), & \dfrac{\partial c_3}{\partial x_3} \end{matrix} \right\|. \tag{58}$$

If we have deformation of a continuous medium and $\overline{MC}$ is the displacement vector, i.e. the vector giving the displacement of a point M of the medium, matrix (58) defines the so-called *deformation tensor*. The anti-symmetric part is:

$$\left\| \begin{matrix} 0, & \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_2} - \dfrac{\partial c_2}{\partial x_1}\right), & \dfrac{1}{2}\left(\dfrac{\partial c_1}{\partial x_3} - \dfrac{\partial c_3}{\partial x_1}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial c_2}{\partial x_1} - \dfrac{\partial c_1}{\partial x_2}\right), & 0, & \dfrac{1}{2}\left(\dfrac{\partial c_2}{\partial x_3} - \dfrac{\partial c_3}{\partial x_2}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial c_3}{\partial x_1} - \dfrac{\partial c_1}{\partial x_3}\right), & \dfrac{1}{2}\left(\dfrac{\partial c_3}{\partial x_2} - \dfrac{\partial c_2}{\partial x_3}\right), & 0 \end{matrix} \right\|. \tag{59}$$

We had an example before of splitting a tensor into two parts, in the particular case of a linear homogeneous deformation [II, 113], when we saw that the anti-symmetric part corresponded to a rotation of the space as a whole (without deformation) about a certain axis.

25. The case of n-dimensional complex space. We now turn to the general case of n-dimensional space. We have already defined a vector in such space as a sequence of n real or complex numbers [12]:

$$\mathbf{x}\,(x_1, x_2, \ldots, x_n),$$

the numbers being referred to as the components of $\mathbf{x}$. We shall assume that the space is referred to a definite fundamental set

$$\mathbf{a}^{(1)}\,(1, 0, \ldots, 0); \quad \mathbf{a}^{(2)}\,(0, 1, \ldots, 0); \quad \ldots; \quad \mathbf{a}^{(n)}\,(0, 0, \ldots, 1),$$

so that

$$\mathbf{x} = x_1\,\mathbf{a}^{(1)} + x_2\,\mathbf{a}^{(2)} + \ldots + x_n\,\mathbf{a}^{(n)}. \tag{60}$$

Vector equality and the elementary operations on vectors were defined by us in [12].

We define a linear transformation of n-dimensional space as the passage from $\mathbf{x}(x_1, x_2, \ldots, x_n)$ to $\mathbf{y}(y_1, y_2, \ldots, y_n)$ in accordance with the formulae:

$$y_i = a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n \qquad (i = 1, 2, \ldots, n), \tag{61}$$

or alternatively:

$$\mathbf{y} = A\mathbf{x}, \tag{62}$$

where A is the matrix $\| a_{ik} \|_1^n$ of the transformation. If its determinant $D(A)$ differs from zero, transformation (62) is said to be non-singular, and A is a non-singular matrix. In this case, on solving equations (61) with respect to the x_i, we get the inverse transformation to (61) or (62):

$$\mathbf{x} = A^{-1} \mathbf{y}, \tag{63}$$

where the matrix A^{-1} has elements

$$\{A^{-1}\}_{ik} = \frac{A_{ki}}{D(A)}, \tag{64}$$

$D(A)$ being the determinant of matrix A and A_{ik} the cofactor of its element a_{ik}.

The definition of the product of two transformations is also analogous to the previous definition [21]: successive application of the two transformations

$$\mathbf{y} = A\mathbf{x}; \quad \mathbf{z} = B\mathbf{y}$$

is equivalent to the single linear transformation

$$\mathbf{z} = BA\mathbf{x}$$

which is called the product of transformations A and B, and the matrix of which is given by

$$\{BA\}_{ik} = \sum_{s=1}^{n} \{B\}_{is} \{A\}_{sk}. \tag{65}$$

The product in general depends on the order of the factors, i.e. we have, apart from exceptional cases:

$$BA \neq AB.$$

The definition of product is readily extended to the case of any number of factors, the associative law being applicable here, i.e. factors may be grouped:

$$(CB) A = C (BA). \tag{66}$$

The inverse transformation satisfies the relationships:

$$AA^{-1} = A^{-1}A = I; \quad (A^{-1})^{-1} = A\,, \tag{67}$$

where I is the so-called unit matrix whose elements are unity on the principal diagonal and zero elsewhere. The unit matrix corresponds to the identity transformation

$$y_i = x_i \quad (i = 1, 2, \ldots, n)\,.$$

We define a diagonal matrix of order n as above:

$$[k_1, k_2, \ldots, k_n] = \begin{Vmatrix} k_1, & 0, & 0, & \ldots, & 0 \\ 0, & k_2, & 0, & \ldots, & 0 \\ 0, & 0, & k_3, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \ldots, & k_n \end{Vmatrix}. \tag{68}$$

This corresponds to the transformation:

$$y_i = k_i x_i \quad (i = 1, 2, \ldots, n)\,.$$

The product of diagonal matrices is independent of the order of the factors and is given by

$$\begin{aligned}[k_1, k_2, \ldots, k_n]\;[l_1, l_2, \ldots, l_n] &= [l_1, l_2, \ldots, l_n]\;[k_1, k_2, \ldots, k_n] \\ &= [k_1 l_1, k_2 l_2, \ldots, k_n l_n]\,.\end{aligned}$$

In the particular case, $k_1 = k_2 = \ldots = k_n = k$, we get the matrix

$$[k, k, \ldots, k] = \begin{Vmatrix} k, & 0, & 0, & \ldots, & 0 \\ 0, & k, & 0, & \ldots, & 0 \\ 0, & 0, & k, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \ldots, & k \end{Vmatrix}, \tag{69}$$

corresponding to multiplication of all the components of a vector by the number k. We shall take matrix (69) for the simple number k, i.e. a number as a particular case of a matrix, which corresponds to what we said towards the beginning of this article. It may easily be seen by using (65) that the product of a number k, treated as matrix (69), with any matrix A is independent of the order of the factors and reduces to multiplication of all the elements of A by the number k:

$$\{[k, k, \ldots, k]\,A\}_{ik} = \{kA\}_{ik} = k\{A\}_{ik}\,. \tag{70}$$

Now suppose we have taken a new fundamental set $\mathbf{b}^{(k)}$ in place of the $\mathbf{a}^{(k)}$ above, the new set being expressed in terms of the $\mathbf{a}^{(k)}$ by the formulae:

$$\left.\begin{array}{l} \mathbf{b}^{(1)} = t_{11}\mathbf{a}^{(1)} + t_{12}\mathbf{a}^{(2)} + \dots + t_{1n}\mathbf{a}^{(n)} \\ \mathbf{b}^{(2)} = t_{21}\mathbf{a}^{(1)} + t_{22}\mathbf{a}^{(2)} + \dots + t_{2n}\mathbf{a}^{(n)} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \mathbf{b}^{(n)} = t_{n1}\mathbf{a}^{(1)} + t_{n2}\mathbf{a}^{(2)} + \dots + t_{nn}\mathbf{a}^{(n)}, \end{array}\right\} \tag{71}$$

where the determinant made up of elements t_{ik} does not vanish. We can now, conversely, write the $\mathbf{a}^{(k)}$ linearly in terms of the $\mathbf{b}^{(k)}$, and any linear combination of the $\mathbf{a}^{(k)}$ is at the same time a linear combination of $\mathbf{b}^{(k)}$, and vice versa. In other words, the $\mathbf{b}^{(k)}$ taken as the fundamental set form the same space as the $\mathbf{a}^{(k)}$. If a vector $\mathbf{x}$ has components $(x_1, \dots, x_n)$ in the system of coordinates defined by the fundamental set $\mathbf{a}^{(k)}$, in the system defined by the fundamental set $\mathbf{b}^{(k)}$ it will have different components $(x_1', \dots, x_n')$, these being expressible in terms of the previous components by means of the linear transformation contragredient to transformation (71), which can be written as

$$(x_1', \dots, x_n') = T^{(*)-1}(x_1, \dots, x_n), \tag{72}$$

where $T^{(*)}$ is the transpose of matrix T of (71).

If we have a spatial transformation given by (62) in the original coordinate system, it will be given in the new coordinate system by

$$\mathbf{y}' = UAU^{-1}\mathbf{x}', \tag{73}$$

where

$$U = T^{(*)-1}.$$

The matrix

$$UAU^{-1}$$

is said to be similar to A.

The basic concepts in the above discussion are those of vector and matrix. Sometimes the vector $\mathbf{x}(x_1, \dots, x_n)$ is itself regarded as a matrix, where one column, no matter which, consists of the numbers $x_1, \dots, x_n$, whilst the remaining columns are filled with zeros. Suppose, for instance, that we put the vector components in the first column; then the vector becomes in matrix form:

$$\begin{Vmatrix} x_1, & 0, & \dots, & 0 \\ x_2, & 0, & \dots, & 0 \\ \dots & & & \\ x_n, & 0, & \dots, & 0 \end{Vmatrix}.$$

Such a matrix in which only the elements of one column are non-zero, is occasionally denoted by

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{Vmatrix} x_1 & 0, & \dots & 0 \\ x_2 & 0, & \dots, & 0 \\ \cdot & \cdot & \cdot & \cdot \\ x_n, & 0, & \dots, & 0 \end{Vmatrix}. \tag{74}$$

We now show that linear transformation (62) can be written as the product of matrix (74) and the transformation matrix A. We multiply matrix (74) by A in accordance with rule (65) and use the fact that only the elements of the first column of (74) are non-zero; this means that only the elements of the first column of the matrix product differ from zero, whilst it is easily seen that the non-zero elements are

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n,$$

i.e. they in fact give linear transformation (62). We can thus write (62) in the form:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \tag{75}$$

where the right-hand side is the product of two matrices.

We conclude the present section by noticing again the general laws obeyed by operations on vectors in n-dimensional space:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}; \quad (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}).$$

If $\mathbf{x}$ and $\mathbf{y}$ are any two vectors, the vector $\mathbf{z} = \mathbf{y} - \mathbf{x}$ is unique, has components $(y_k - x_k)$, and satisfies $\mathbf{x} + \mathbf{z} = \mathbf{y}$.

Let a and b be any numbers. We have:

$$(a + b)\,\mathbf{x} = a\mathbf{x} + b\mathbf{x}; \quad a\,(b\mathbf{x}) = (ab)\,\mathbf{x}; \quad a\,(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}.$$

We have $1\mathbf{x} = \mathbf{x}$ for the number unity, and $0\mathbf{x} = 0$, where the 0 on the right-hand side denotes the vector, all the components of which vanish.

26. Basic matrix calculus. Earlier sections have contained formulae in which the matrix appears as a new symbol, on which a number of operations analogous to those on ordinary numbers could be carried out. This naturally suggests the idea of constructing a new algebra which would be satisfied by symbols representing matrices. In other

words, we propose to regard a matrix as a new type of number, or as a *hypercomplex* number. Just as we arrived previously with the aid of two real numbers at the idea of a complex number of the form $a + ib$, so now we arrive at the new number, the *matrix*, with the aid of n^2 complex numbers a_{ik}, arranged as a square array. An essential difference must be pointed out, however. We have seen that all the formal operations of the algebra of real numbers can be carried out on the letters symbolizing complex numbers. The same cannot be said as regards matrices. Matrix algebra derives its fundamental difference from the algebra of complex numbers from the *non-commutative nature of multiplication*, i.e. a product depends on the order of the factors. We now propose to establish the basic rules of matrix algebra; the results already obtained when regarding a matrix as the array of a linear transformation will be used as a guide in most relationships.

We shall consider square matrices of the same order n throughout what follows, unless some special remark is made. We use the same notation as above, of $\{A\}_{ik}$ for the elements of the matrix A.

Two matrices A and B are reckoned equal when, and only when,

$$\{A\}_{ik} = \{B\}_{ik} \qquad (i, k, = 1, 2, \ldots, n), \tag{76}$$

i.e. when all corresponding elements are the same.

A matrix sum is defined by the formula:

$$\{A + B\}_{ik} = \{A\}_{ik} + \{B\}_{ik}, \tag{77}$$

i.e. corresponding elements are added.

Multiplication is defined by

$$\{BA\}_{ik} = \sum_{s=1}^{n} \{B\}_{is} \{A\}_{sk}. \tag{78}$$

As we saw above, in general

$$BA \neq AB,$$

though the associative law is valid [21]:

$$(CB)A = C(BA). \tag{79}$$

The determinant of a product is equal to the product of the determinants of the matrix factors:

$$D(BA) = D(B) \cdot D(A). \tag{80}$$

The distributive law is clearly also valid:

$$(A + B)C = AC + BC \quad \text{and} \quad C(A + B) = CA + CB. \tag{81}$$

A special feature of matrix multiplication should be noted: a product can be zero, i.e. all its elements can vanish, although none of the individual factors vanish. The example may be quoted of the two identical second order matrices:

$$\begin{Vmatrix} 0, & 0 \\ 1, & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 0, & 0 \\ 1, & 0 \end{Vmatrix} = \begin{Vmatrix} 0, & 0 \\ 0, & 0 \end{Vmatrix}.$$

The concept of the inverse matrix A^{-1} is brought in exactly as in a previous section; A must be non-singular, i.e. $D(A) \neq 0$. If $C = BA$ and R_A, R_B, R_C are the ranks of A, B, C, we saw in [7] that $R_C \leqslant R_A$. If B is non-singular, $A = B^{-1}C$, and we can say as above that $R_A \leqslant R_C$, and consequently, $R_C = R_A$, i.e. *the rank of a matrix A is unchanged on multiplication on the right or left by a non-singular matrix B*. We have the relationships for the unit matrix I:

$$BI = IB = B, \tag{82}$$

where B is any matrix.

It may easily be seen that A^{-1} is the unique solution of the equations

$$AX = I \quad \text{and} \quad XA = I, \tag{83}$$

where I is the unit matrix. For, on multiplying say the first equation on the left by A^{-1} and taking (79) and (67) into account, we get $X = A^{-1}$, and similarly for the second equation. It must be noticed that (83) has no solution whatever if $D(A) = 0$, i.e. no inverse of A exists. For otherwise, (83) would give us

$$D(A)\,D(X) = 1,$$

which contradicts the condition $D(A) = 0$.

The concept of diagonal matrix should be recalled from the previous section, as also the fact that any number k can be regarded as a particular case of a matrix. We can easily bring in positive integral powers of a matrix:

$$A^p = A \cdot A \ldots A.$$

Negative integral powers of a matrix are introduced as positive integral powers of the inverse matrix, i.e

$$A^{-p} = (A^{-1})^p. \tag{84}$$

We obviously have

$$A^{-p} = (A^p)^{-1}, \quad \text{i.e.} \quad A^{-p}A^p = A^pA^{-p} = I. \tag{85}$$

The symbol for the quotient of two matrices:

$$\frac{A}{B}$$

does not have a definite meaning. We can interpret it in two ways: as the product AB^{-1}, or as $B^{-1}A$; these products are in general distinct, and it is only in the particular cases when they coincide that the quotient symbol has an exact significance.

A further basic concept is that of similar matrices which we also introduced in the previous section. We shall note a number of formulae which are very easily proved:

$$(CBA)^{-1} = A^{-1} B^{-1} C^{-1}, \tag{86}$$

$$CBAC^{-1} = (CBC^{-1})(CAC^{-1}). \tag{87}$$

If $A^{(*)}$ denotes the transpose of any matrix A, we also have:

$$(CBA)^{(*)} = A^{(*)} B^{(*)} C^{(*)}, \tag{88}$$

which is easily verified by using the definition of product. Two new notations must be introduced. We shall write $\bar{A}$ for the matrix whose elements are the conjugates of the elements of A, i.e.

$$\{\bar{A}\}_{ik} = \overline{\{A\}}_{ik}, \tag{89}$$

a symbol of the type $\bar{a}$ being used as usual to denote the conjugate of a. Lastly, we shall write $\tilde{A}$ for the matrix obtained by interchanging rows and columns in A and replacing all the elements by their conjugates, i.e.

$$\{\tilde{A}\}_{ik} = \overline{\{A\}}_{ki}. \tag{90}$$

The matrix $\tilde{A}$ is generally called the Hermitian or Hermitian conjugate of matrix A (due to Hermite, a French mathematician of the latter half of the nineteenth century). It may easily be verified that

$$\widetilde{CBA} = \tilde{A}\tilde{B}\tilde{C}. \tag{91}$$

We suggest that the following elementary formula also be verified:

$$(A^{(*)})^{-1} = A^{-1(*)},$$

i.e. the signs of the inverse and transpose can change places, as already mentioned in [20].

We notice an expression which will be useful later. It follows at once from (67) that

$$D(A)D(A^{-1}) = 1,$$

i.e.

$$D(A^{-1}) = D(A)^{-1}. \tag{92}$$

In other words, the determinant of the inverse matrix has a value equal to the reciprocal of the determinant of the original matrix.

The concept of diagonal matrix may be generalized to that of *quasi-diagonal matrix*. This will be explained in a particular case. Let us take the seventh order matrix:

$$\left\| \begin{matrix} b_{11}, & b_{12}, & b_{13}, & 0, & 0, & 0, & 0 \\ b_{21}, & b_{22}, & b_{23}, & 0, & 0, & 0, & 0 \\ b_{31}, & b_{32}, & b_{33}, & 0, & 0, & 0, & 0 \\ 0, & 0, & 0, & c_{11}, & c_{12}, & 0, & 0 \\ 0, & 0, & 0, & c_{11}, & c_{22}, & 0, & 0 \\ 0, & 0, & 0, & 0, & 0, & d_{11}, & d_{12} \\ 0, & 0, & 0, & 0, & 0, & d_{22}, & d_{22} \end{matrix} \right\|.$$

Let B denote the third order matrix with elements b_{ik}, and C and D the second order matrices with elements c_{ik} and d_{ik}, respectively. The above seventh order matrix is called a *quasi-diagonal* $\{3, 2, 2\}$ structure and is denoted by

$$[B, C, D].$$

We suppose in general that the principal diagonal of an nth order matrix, made up of the elements a_{ii}, is divided into m parts, the first part consisting of the first k_1 elements, the second of the next k_2 elements, and so on, so that $k_1 + \ldots + k_m = n$. We can regard the first k_1 elements as the principal diagonal of a matrix X_1 of order k_1; the next k_2 elements as the principal diagonal of a matrix X_2 of order k_2, and so on. Suppose that all the elements of our matrix A not belonging to matrices X_s are zero. Then A is known as a quasi-diagonal matrix of structure $\{k_1, \ldots, k_m\}$ and is written symbolically:

$$A = [X_1, X_2, \ldots, X_m].$$

The rules for operating on quasi-diagonal matrices of the same structure are of unusual simplicity. We shall state the relevant formulae and omit the proofs: these are based on the definitions of the operations and are purely elementary. We have for addition of quasi-diagonal matrices of the same structure:

$$\begin{aligned}[X_1, X_2, \ldots, X_m] + [Y_1, Y_2, \ldots, Y_m] = \\ = [X_1 + Y_1, X_2 + Y_2, \ldots, X_m + Y_m],\end{aligned} \tag{93}$$

where the fact of the same structure implies that every matrix X_k is of the same order as the corresponding matrix Y_k. Similarly, we

have for multiplication, and for raising to a power:

$$[Y_1, Y_2, \dots, Y_m]\,[X_1, X_2, \dots, X_m] = [Y_1 X_1, Y_2 X_2, \dots, Y_m X_m], \tag{94}$$

$$[X_1, X_2, \dots, X_m]^p = [X_1^p, X_2^p, \dots, X_m^p], \tag{95}$$

where p is any positive or negative integer, except, of course, that none of the determinants $D(X_k)$ must vanish if p is negative.

The rule for the similarity transformation of a matrix $[X_1, X_2, \dots, X_m]$ with the aid of a matrix of the same structure is given by

$$[Y_1, \dots, Y_m]\,[X_1, \dots, X_m]\,[Y_1, \dots, Y_m]^{-1} =$$
$$= [Y_1 X_1, Y_1^{-1}, \dots, Y_m X_m Y_m^{-1}]. \tag{96}$$

The geometric interpretation of the linear transformations supplied by quasi-diagonal matrices is seen as follows. We take for simplicity the seventh order matrix used above, the structure of which is defined by the numbers $\{3, 2, 2\}$, and we consider the corresponding linear transformation. If we have in the original vector $(x_1, \dots, x_7)$:

$$x_4 = x_5 = x_6 = x_7 = 0,$$

obviously, in the transformed vector:

$$y_4 = y_5 = y_6 = y_7 = 0.$$

so that, in fact, all the vectors belonging to the subspace formed by the first three fundamental vectors belong to the same subspace after transformation, and the transformation will itself be defined by a third order matrix B. The same applies to the subspace formed by the next two fundamental vectors, and similarly, to that formed by the last two.

It may be recalled here that the *subspace formed by vectors* $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(l)}$ is defined as the system of vectors given by

$$c_1 \mathbf{x}^{(1)} + \dots + c_l \mathbf{x}^{(l)},$$

where $c_1, \dots, c_l$ are arbitrary constants.

27. Characteristic roots of matrices and reduction to canonical form. Though similar matrices are obviously not equal in the sense of (76), they are equivalent in the geometrical sense inasmuch as they embody the same linear transformation of space, expressed in different coordinate systems. We shall now look for the invariants of these matrices, i.e. the expressions made up of elements which would have the same values for all similar matrices. One invariant is easily found.

This is the determinant of the matrix. For, given A, let UAU^{-1} be a similar matrix, where U is any matrix with non-zero determinant. We have by (80) and (92):

$$D(UAU^{-1}) = D(U)\,D(A)\,D(U^{-1}) = D(U)\,D(A)\,D(U)^{-1} = D(A).$$

We form another invariant by taking the polynomial $\varphi(\lambda)$ of degree n in a parameter λ, equal to the determinant of the matrix obtained from A by subtracting λ from each diagonal element, i.e. by taking

$$\varphi(\lambda) = \begin{vmatrix} a_{11} - \lambda, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22} - \lambda, & \ldots, & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} - \lambda \end{vmatrix}, \tag{97}$$

where the a_{ik} are the elements of A. We can write this alternatively as:

$$\varphi(\lambda) = D(A - \lambda) = D(A - \lambda I), \tag{98}$$

since λ or λI is a diagonal matrix by hypothesis, in which all the elements on the principal diagonal are equal to λ. On replacing A by UAU^{-1} and noticing that commutation of λ with any matrix is possible, so that $U\lambda U^{-1} = \lambda$, we have:

$$D(UAU^{-1} - \lambda) = D\left[U(A - \lambda)\,U^{-1}\right] = D(A - \lambda)$$

whence

$$D(UAU^{-1} - \lambda) = D(A - \lambda). \tag{99}$$

Hence we see that polynomial (97) is the same when formed for UAU^{-1} as when formed for A. In other words, all the coefficients of (97) are invariants with respect to a similar matrix. The final coefficient is obviously $(-1)^n$. We shall pay particular attention to the constant term and the coefficient of $(-1)^{n-1}\lambda^{n-1}$. The former is clearly the determinant, an invariant that we have already mentioned. The latter may be seen, on using the results of [5], to be equal to the sum of the diagonal elements. This sum is generally called the *trace of the matrix,* and is written as follows:

$$\mathrm{Tr}(A) = \{A\}_{11} + \{A\}_{22} + \ldots + \{A\}_{nn} = a_{11} + a_{22} + \ldots + a_{nn},$$

where Tr is occasionally written Sp, from the German "Spur" (meaning "trace"). Similar matrices thus have the same determinant and the same trace.

We now write the equation

$$D(A - \lambda) = 0, \tag{100}$$

known as the *characteristic equation of matrix* A, its roots being called the *characteristic roots, or eigenvalues, of* A. We can say from the above that similar matrices have the same characteristic roots. An equation of the form (100) has already been encountered.

We now pose the question: is there a matrix V such that, on carrying out the similarity transformation on a given matrix A, the resulting $V^{-1}AV$ is a diagonal matrix? Or alternatively, expressed from the point of view of linear transformations of space, is it possible to choose coordinate axes such that a linear transformation characterized by the matrix A in the original coordinate system reduces in the new system simply to a transformation of the form $y_k = \lambda_k x_k$? We should remark that the fact of writing $V^{-1}AV$ instead of the previous UAU^{-1} is of no real consequence.

We can write down our condition as:

$$V^{-1}AV = [\lambda_1, \lambda_2, \ldots, \lambda_n], \tag{101}$$

where it is required to find the elements of V and the numbers λ_k. By multiplying both sides on the left by V, we can evidently rewrite the condition as

$$AV = V[\lambda_1, \lambda_2, \ldots, \lambda_n]. \tag{102}$$

We now use (65) to find the elements on both sides with subscripts i and k. This gives us n^2 equations:

$$\sum_{s=1}^{n} a_{is} v_{sk} = v_{ik} \lambda_k,$$

where a_{ik} and v_{ik} are the elements of A and V.

We fix the second subscript k and put $i = 1, 2, \ldots, n$ which gives us n equations containing only the number λ_k and the elements $v_{1k}, \ldots, v_{nk}$ of the kth column of V:

$$\sum_{s=1}^{n} a_{is} v_{sk} = \lambda_k v_{ik} \quad (i = 1, 2, \ldots, n). \tag{103}$$

If we take the elements $(v_{1k}, \ldots, v_{nk})$ as the components of a vector $\mathbf{v}^{(k)}$, we can write the above set as the single vector equation:

$$A\mathbf{v}^{(k)} = \lambda_k \mathbf{v}^{(k)}. \tag{104}$$

Hence the discovery of the matrix V that reduces A to the diagonal form amounts to finding the vectors $\mathbf{v}^{(k)}$ that are reproduced identically except for a numerical factor as a result of the linear transformation defined by A. This is the algebraic analogue of the position in present-day quantum theory, according to which Heisenberg's matrix mechanics is in essence equivalent to Schrödinger's wave mechanics. From the former point of view, the basic problem is that of reducing an (infinite) matrix to the diagonal form. As regards wave mechanics, the essential problem here is that of finding vectors (in space with an infinite set of dimensions) such that they are identically reproduced except for a numerical factor as a result of a linear transformation. We have referred to the above discussion as the algebraic analogue inasmuch as the problems are reduced to purely algebraic problems by confining ourselves to space with a finite number of dimensions. The more complex case of space with an infinite set of dimensions requires an essential departure from ordinary algebra and makes use of the apparatus of analysis. All these questions are treated in detail later, though it may be mentioned meantime that applications to physics in the case of a finite number of dimensions require only matrices A of a particular type (Hermitian matrices for which $a_{ik} = \bar{a}_{ik}$) and matrices U likewise of a definite type (unitary matrices, the definition of which is given below). Although we shall consider here the general problem for any finite matrix, we confine ourselves to the statement of final results without full proofs. A full solution will only be given in the case of problems of practical interest.

We turn to the solution of system (103) or (104). This becomes, when written out in full:

$$\left.\begin{aligned} (a_{11} - \lambda_k)\, v_{1k} + a_{12}\, v_{2k} + \dots + a_{1n}\, v_{nk} &= 0 \\ a_{21}\, v_{1k} + (a_{22} - \lambda_k)\, v_{2k} + \dots + a_{2n}\, v_{nk} &= 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots& \\ a_{n1}\, v_{1k} + a_{n2}\, v_{2k} + \dots + (a_{nn} - \lambda_k)\, v_{nk} &= 0\,. \end{aligned}\right\} \qquad (105)$$

The necessary and sufficient condition for obtaining a non-zero solution for $(v_{1k}, \dots, v_{nk})$ is the vanishing of the determinant of the system, i.e. λ_k must be a characteristic roots of A. We shall only treat in detail the case when the characteristic roots are distinct. Let the roots be $\lambda_1, \dots, \lambda_n$. On substituting the first root λ_1 in the coefficients of system (105), we can find from this the elements of the first column of V. We shall not go into the question of how wide is the choice of the v_{il}. We choose a solution of the system in any uniquely

defined manner with the only proviso that it be non-zero. Similarly, on replacing $\lambda_k = \lambda_2$ in the coefficients of (105), we can find the elements of the second column of V, and so on, as far as the nth column. Equations (105) are equivalent to (102), and all we need in order to arrive at the basic equation (101) is the existence of the inverse V^{-1} of V, i.e. non-vanishing of the determinant of V. We prove this latter property by assuming the contrary, i.e. let the determinant vanish. As we know from [12], this is equivalent to the existence of a linear relationship between the vectors $\mathbf{v}^{(k)}$ defined by the columns of V:

$$C_1 \mathbf{v}^{(1)} + \ldots + C_n \mathbf{v}^{(n)} = 0,$$

where not all the coefficients C_k are zero. We apply the transformation defined by matrix A $(n - 1)$ times to both sides of this equation. Using (104), we have the n equations:

$$\begin{aligned} C_1 \mathbf{v}^{(1)} + \ldots + \quad C_n \mathbf{v}^{(n)} &= 0 \\ \lambda_1 C_1 \mathbf{v}^{(1)} + \ldots + \quad \lambda_n C_n \mathbf{v}^{(n)} &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \lambda_1^{n-1} C_1 \mathbf{v}^{(1)} + \ldots + \lambda_n^{n-1} C_n \mathbf{v}^{(n)} &= 0. \end{aligned}$$

Since not all the vectors $C_k \mathbf{v}^{(k)}$ vanish, we can say that the determinant of the system must vanish:

$$\begin{vmatrix} 1, & 1, & \ldots, & 1 \\ \lambda_1, & \lambda_2, & \ldots, & \lambda_n \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1^{n-1}, & \lambda_2^{n-1}, & \ldots, & \lambda_n^{n-1} \end{vmatrix} = 0,$$

where the numbers λ_k are distinct by hypothesis. But this last equation contradicts the fact that the Vandermonde determinant of distinct numbers cannot vanish. We have thus proved the possibility of reducing the matrix by means of the similarity transformation to the diagonal form in the case when all the characteristic roots of the matrix are distinct. When some characteristic roots are equal, it may happen that the matrix cannot be reduced by a similarity transformation to the diagonal form. None the less, there exists in this case a simplest or *canonical* form of the matrix. In the case when the matrix reduces to the diagonal form, the canonical form becomes

$$[\lambda_1\ \lambda_2, \ldots, \lambda_n],$$

where λ_k are the characteristic roots of the matrix. We shall merely state the result in the general case.† Let $\lambda = a$ be a root of equation (100) of multiplicity k. Further, let $\lambda = a$ be a root of multiplicity k_1 but not more than k_1, of all the $(n-1)$th order determinants of the array on the left-hand side of (100), i.e. every such determinant is divisible by $(\lambda - a)^{k_1}$, but at least one is not divisible by $(\lambda - a)^{k_1+1}$. Similarly, let all the $(n-2)$th order determinants have $\lambda = a$ as a root of multiplicity k_2, but not more than k_2, and so on, till finally the same root is of multiplicity k_m for all the $(n-m)$th order determinants, whereas at least one of the $(n-m-1)$th order determinants is non-vanishing for $\lambda = a$. This last will evidently be true for the successive lower order determinants. It can be shown that the sequence of numbers k_s is decreasing, i.e.

$$k > k_1 > k_2 > \ldots > k_m.$$

We bring in the following positive integers:

$$l_1 = k - k_1; \quad l_2 = k_1 - k_2; \ldots; \quad l_{m+1} = k_m,$$

where obviously, $l_1 + l_2 + \ldots + l_{m+1} = k$.

The expressions:

$$(\lambda - a)^{l_1}; \quad (\lambda - a)^{l_2}; \ldots; \quad (\lambda - a)^{l_{m+1}}$$

are known as the *elementary divisors of matrix A* corresponding to the root $\lambda = a$. We can similarly find the elementary divisors for all other characteristic roots of A and hence obtain the *set of elementary divisors:*

$$(\lambda - \lambda_1)^{\varrho_1}; \quad (\lambda - \lambda_2)^{\varrho_2}; \quad \ldots; \quad (\lambda - \lambda_p)^{\varrho_p}, \tag{106}$$

where

$$\varrho_1 + \varrho_2 + \ldots + \varrho_p = n \tag{107}$$

and not all the λ_k need be distinct.

We saw above that the characteristic roots are unchanged by a similarity transformation. The set of elementary divisors of a matrix happens to possess the same property. We now introduce some new elementary matrices $I_\varrho(a)$, where the symbol represents the matrix of order ϱ in which a is repeated down the main diagonal, unity is

† A proof will be found in a special note at the end of Part 2 of this Volume.

repeated down the diagonal below, and the remaining elements are zero:

$$I_\varrho(a) = \begin{Vmatrix} a, & 0, & 0, & \dots, & 0, & 0 \\ 1, & a, & 0, & \dots, & 0, & 0 \\ 0, & 1, & a, & \dots, & 0, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \dots, & a, & 0 \\ 0, & 0, & 0, & \dots, & 1, & a \end{Vmatrix}. \tag{108}$$

The following result is fundamental for the problem of representing a matrix in the canonical form: if A has elementary divisors (106), there exists a matrix U with a non-zero determinant such that

$$UAU^{-1} = [I_{\varrho_1}(\lambda_1),\ I_{\varrho_2}(\lambda_2),\ \dots,\ I_{\varrho_p}(\lambda_p)]. \tag{109}$$

We mention that finding U reduces to elementary algebraic operations if all the characteristic roots of A are known. If $\varrho = 1$, $I_\varrho(a)$ is understood to mean simply the number a. It can happen that, even with the existence of equal characteristic roots, all the elementary divisors (106) are simple, i.e. have the form

$$(\lambda - \lambda_1);\quad (\lambda - \lambda_2);\quad \dots;\quad (\lambda - \lambda_n).$$

In this case the quasi-diagonal matrix

$$[I_{\varrho_1}(\lambda_1),\ I_{\varrho_2}(\lambda_2),\ \dots,\ I_{\varrho_p}(\lambda_p)]$$

reduces simply to the diagonal matrix $[\lambda_1, \dots, \lambda_n]$, and we have reduction of the matrix to the diagonal form.

It must be pointed out that the matrix U appearing in (109) is not uniquely defined. In particular, if d is the magnitude of the determinant of U, we can replace

$$U \text{ by } \frac{1}{\sqrt[n]{d}}U \quad \text{and} \quad U^{-1} \text{ by } \sqrt[n]{d}\,U,$$

in (109), and hence we can take the determinant of U in (109) as equal to unity. Our treatment of the general problem of reducing a matrix to the canonical form is limited to these remarks for the present, though we return to discuss it in a special note at the end of Part 2 of Vol. III. As already mentioned, a detailed treatment of the problem for a particular type of matrix will be found below.

It is easily shown that the necessary and sufficient condition for a matrix to be reducible to the diagonal form is for the rank of the matrix

of the coefficients of system (105) to be equal to $(n - \mu_k)$, where μ_k is the multiplicity of the root λ_k of the secular equation. When this condition is satisfied, system (105) defines μ_k linearly independent vectors $(v_{1k}, v_{2k}, \ldots, v_{nk})$ [14].

28. Unitary and orthogonal transformations. We shall make use in this and later sections of the concepts of scalar product and vector norm (length), introduced in [13]. We recall that the square of the norm (length) is defined by

$$\|\mathbf{x}\|^2 = (\mathbf{x}, \mathbf{x}) = \sum_{s=1}^{n} |x_s|^2, \tag{110}$$

or, in the case of real components:

$$\|\mathbf{x}\|^2 = \sum_{s=1}^{n} x_s^2.$$

This definition of norm is bound up with a definite choice of fundamental vectors, i.e. of coordinate axes. We shall refer to the coordinate system in the above definition of norm as a normal or Cartesian system. Apart form vector length, we have defined the *scalar product of two vectors* by the formula

$$(\mathbf{x}, \mathbf{y}) = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n. \tag{111}$$

In the case of real vectors, this expression takes the more symmetrical form

$$(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

It follows from (111) that the scalar product changes to its conjugate on changing the order of the vectors:

$$(\mathbf{y}, \mathbf{x}) = \overline{(\mathbf{x}, \mathbf{y})}. \tag{112}$$

We have described two vectors as perpendicular or orthogonal when their scalar product is zero.

In future, unless something is said specifically, we assume that a Cartesian coordinate system is in question. In view of this, the linear transformations corresponding to passage from one Cartesian system to another have a special significance. We know that there is a corresponding linear transformation of components for every passage from one fundamental vector set to another. Let us take the transformation

$$(y_1, \ldots, y_n) = U(x_1, \ldots, x_n), \tag{113}$$

where the original coordinate system is Cartesian. The necessary and sufficient condition for the new system to be Cartesian is for the length of a vector in the new system to be likewise given by the sum of the squares of the moduli of the components, i.e.

$$|y_1|^2 + \dots + |y_n|^2 = |x_1|^2 + \dots + |x_n|^2. \tag{114}$$

We show that, with this, the value of a scalar product is given in the new system by an expression analogous to (111). For suppose we have the two vectors

$$\mathbf{x}\,(x_1 \dots, x_n) \text{ and } \mathbf{x}'\,(x_1', \dots, x_n'),$$

in the original coordinate system, corresponding in the new system to

$$\mathbf{y}\,(y_1 \dots, y_n) \text{ and } \mathbf{y}'\,(y_1', \dots, y_n').$$

We form two new vectors: $\mathbf{z} = \mathbf{x} + \mathbf{x}'$ and $\mathbf{u} = \mathbf{x} + i\mathbf{x}'$, with components $(x_k + x_k')$ and $(x_k + ix_k')$. Assuming condition (114) fulfilled, we have

$$\sum_{k=1}^{n} (y_k + y_k')\,(\overline{y}_k + \overline{y}_k') = \sum_{k=1}^{n} (x_k + x_k')\,(\overline{x}_k + \overline{x}_k'),$$

whence, again by (114), we get finally:

$$\sum_{k=1}^{n} (y_k \overline{y}_k' + y_k' \overline{y}_k) = \sum_{k=1}^{n} (x_k \overline{x}_k' + x_k' \overline{x}_k), \tag{115$_1$}$$

since

$$\sum_{k=1}^{n} |y_k|^2 = \sum_{k=1}^{n} |x_k|^2 \text{ and } \sum_{k=1}^{n} |y_k'|^2 = \sum_{k=1}^{n} |x_k'|^2.$$

Similarly:

$$\sum_{k=1}^{n} (y_k + iy_k')\,(\overline{y}_k - i\overline{y}_k') = \sum_{k=1}^{n} (x_k + ix_k')\,(\overline{x}_k - i\overline{x}_k')$$

and consequently:

$$\sum_{k=1}^{n} (y_k' \overline{y}_k - y_k \overline{y}_k') = \sum_{k=1}^{n} (x_k' \overline{x}_k - x_k \overline{x}_k'). \tag{115$_2$}$$

Equations (115_1) and (115_2) give:

$$\sum_{k=1}^{n} y_k \overline{y}_k' = \sum_{k=1}^{n} x_k \overline{x}_k', \tag{116}$$

i.e. the scalar product is in fact given by the previous formula. Hence,

if transformation (113) satisfies condition (114), it likewise satisfies (116), i.e. the value of the scalar product remains invariant. Conversely, (114) follows from (116), if we put $x'_k = x_k$ in (116), since the scalar product of two identical vectors obviously reduces to the square of the length of the vector. Linear transformations that satisfy condition (114) or (116) are generally called *unitary*.

If we take real space and linear transformations with real matrices, condition (114) becomes simply

$$y_1^2 + y_2^2 + \ldots + y_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2, \tag{117}$$

and the corresponding real transformations are called *orthogonal*. They are evidently particular cases of unitary transformations.

We shall now elucidate the special properties of unitary transformations. We write conditions (114) for transformation (113) in the explicit form, the elements of U being denoted by u_{ik}:

$$\sum_{k=1}^{n} | u_{k1} x_1 + \ldots + u_{kn} x_n |^2 = \sum_{k=1}^{n} | x_k |^2$$

or

$$\sum_{k=1}^{n} (u_{k1} x_1 + \ldots + u_{kn} x_n)(\bar{u}_{k1} \bar{x}_1 + \ldots + \bar{u}_{kn} \bar{x}_n) = \sum_{k=1}^{n} x_k \bar{x}_k. \tag{118}$$

On removing the brackets on the left-hand side and equating coefficients of $x_p \bar{x}_p$ to unity, and of $x_p \bar{x}_q$ $(p \neq q)$ to zero, we have the necessary and sufficient condition for the elements of a unitary transformation in the following form:

$$\left.\begin{aligned} &\sum_{k=1}^{n} | u_{kp} |^2 = 1 \qquad (p = 1, 2, \ldots, n), \\ &\sum_{k=1}^{n} u_{kp} \bar{u}_{kq} = 0 \qquad (p \neq q), \end{aligned}\right\} \tag{119}$$

i.e. the sum of the squares of the moduli of the elements of each column must be equal to unity and the sum of the products of element of one column with the conjugates of corresponding elements of another column must be zero. These conditions are sometimes written as

$$\sum_{k=1}^{n} u_{kp} \bar{u}_{kq} = \delta_{pq}, \tag{120}$$

where δ_{pq} are the elements of the unit matrix, i.e.

$$\delta_{pq} = \begin{cases} 0 & (p \neq q) \\ 1 & (p = q) \end{cases}. \tag{121}$$

We applied above to identity (118) the method of undetermined coefficients. This is sufficient, of course, for satisfying the identity. It is easy to show, by assigning particular values to x_k, that identity of the coefficients of like terms is also necessary.

We take determinants $D(A)$ and $D(\bar{A})$, the latter consisting of conjugate elements of the former. On multiplying these column by column [6], we obtain by (119) the determinant of the unit matrix, i.e. unity. On the other hand, it is clear that both our determinants are expressed by complex conjugate numbers, and it follows at once from what has been said that

$$|D(A)|^2 = 1,$$

i.e. *the square of the modulus of the determinant of a unitary matrix is equal to unity.* In other words, the determinant of a unitary matrix has a modulus of unity, i.e. is expressed by a complex number of the form $e^{i\varphi}$, where φ is real.

We introduce into the discussion $U^{(*)}$, the transpose of matrix U.

Conditions (119), which are generally known as column orthogonality conditions, can be written in the matrix form

$$\bar{U}^{(*)} U = I, \tag{122}$$

which is equivalent to

$$U^{-1} = \bar{U}^{(*)} = \tilde{U}, \tag{123}$$

i.e. if a matrix is unitary, its inverse is equal to its Hermitian conjugate.

The transformation U^{-1}, the inverse of U, expresses the passage from vectors $\mathbf{y}$ to $\mathbf{x}$. This also clearly satisfies unitary condition (114), i.e. if U is a unitary matrix, its inverse U^{-1} is also unitary. In other words, by (123), $\tilde{U}$ is unitary, and its columns satisfy the orthogonality conditions. But the columns of $\tilde{U}$ are the rows of $\bar{U}$. We can thus say that the rows as well as the columns of a unitary matrix satisfy orthogonality conditions, i.e. we have in addition to (120):

$$\sum_{k=1}^{n} u_{pk}\, \bar{u}_{qk} = \delta_{pq}. \tag{124}$$

Similarly to the above, if matrices U_1 and U_2 satisfy condition (114), their product $U_1 U_2$ also clearly satisfies this condition, i.e. the product of two unitary matrices is also unitary.

We indicate two alternative forms of the definition of unitary matrix:

$$|U\mathbf{x}|^2 = |\mathbf{x}|^2 \text{ or } (U\mathbf{x}, U\mathbf{x}') = (\mathbf{x}, \mathbf{x}'), \tag{125_1}$$

$\mathbf{x}$ and $\mathbf{x}'$ in the second equation being arbitrary vectors.

We now consider the situation when a unitary matrix has real elements. As already mentioned, it is described as orthogonal in this case, the corresponding transformation being an orthogonal transformation. Here we have, instead of (120) and (124):

$$\sum_{k=1}^{n} u_{kp} u_{kq} = \delta_{pq}; \qquad \sum_{k=1}^{n} u_{pk} u_{qk} = \delta_{pq}. \tag{125$_2$}$$

Moreover the determinant of the transformation must certainly be a real number, so that its value can only be ± 1. These real orthogonal transformations in n-dimensional space are the complete analogue of the transformations of three-dimensional space that we discussed in [20]. In the real case, moreover, $\tilde{U}$ coincides with $U^{(*)}$, i.e. the inverse transformation U^{-1} is got from U by replacing rows by columns.

We mention further that every complex number $e^{i\varphi}$, where φ is real, regarded as the matrix $[e^{i\varphi}, e^{i\varphi}, \ldots, e^{i\varphi}]$, represents a unitary matrix, and if U is a unitary matrix, the product $e^{i\varphi} U$ is likewise unitary. We explained in [25] the meaning of the product of a number with a matrix.

29. Buniakowski's inequality. In the present section we establish an inequality which will be useful later. We have already derived this inequality in Vol. II [II, 156]. It consists in the following: whatever the real numbers $a_1, a_2, \ldots, a_m$ and $\beta_1, \beta_2, \ldots, \beta_m$, we have:

$$\left(\sum_{k=1}^{m} a_k \beta_k\right)^2 \leqslant \sum_{k=1}^{m} a_k^2 \cdot \sum_{k=1}^{m} \beta_k^2, \tag{127}$$

where we have the equals sign when and only when the a_k and β_k are proportional, i.e.

$$\frac{\beta_1}{a_1} = \frac{\beta_2}{a_2} = \ldots = \frac{\beta_m}{a_m}. \tag{127}$$

Let ξ be any real number. We form the sum:

$$S = \sum_{k=1}^{m} (\xi a_k - \beta_k)^2,$$

which is clearly $\geqslant 0$. We have the equality when and only when

$$\frac{\beta_1}{a_1} = \frac{\beta_2}{a_2} = \ldots = \frac{\beta_m}{a_m} = \xi$$

and obviously in this case:

$$\left(\sum_{k=1}^{m} a_k \beta_k\right)^2 = \sum_{k=1}^{m} a_k^2 \cdot \sum_{k=1}^{m} \beta_k^2.$$

Generally speaking, on removing the brackets in S, we get the quadratic expression

$$S = A\xi^2 - 2B\xi + C,$$

where

$$A = \sum_{k=1}^{m} a_k^2; \quad B = \sum_{k=1}^{m} a_k \beta_k; \quad C = \sum_{k=1}^{m} \beta_k^2.$$

The quadratic expression remains $\geqslant 0$ for all real ξ, whence it follows that $AC - B^2 \geqslant 0$, i.e. $B^2 \leqslant AC$, which leads to inequality (126).

If $AC - B^2 = 0$, the quadratic form must vanish for some real ξ, whilst condition (127) must be fulfilled, as we saw. Conversely, if the condition is satisfied, we have the $=$ sign in (126). Now let the a_k and β_k be complex numbers. Using the fact that the modulus of a sum is $\leqslant$ the sum of the moduli of the terms, we get

$$\left| \sum_{k=1}^{m} a_k \beta_k \right| \leqslant \sum_{k=1}^{m} |a_k| |\beta_k| .$$

On applying inequality (126) to the last sum, consisting of positive terms, we get

$$\left| \sum_{k=1}^{m} a_k \beta_k \right|^2 \leqslant \sum_{k=1}^{m} |a_k|^2 \cdot \sum_{k=1}^{m} |\beta_k|^2. \qquad (126_1)$$

It is easily shown that in the present case, with complex a_k and β_k, the sign of equality occurs when and only when $|a_k|$ and $|\beta_k|$ are proportional, and all the products $a_k \beta_k$ have the same argument. Inequality (126) is applicable to integrals as well as sums, as already mentioned in [II, 156]. If $f_1(x)$ and $f_2(x)$ are two real functions in the interval $a \leqslant x \leqslant b$, the inequality for integrals is

$$\left[\int_a^b f_1(x) f_2(x)\, dx \right]^2 \leqslant \int_a^b f_1^2(x)\, dx \cdot \int_a^b f_2^2(x)\, dx. \qquad (126_2)$$

We see this by forming the expression

$$\int_a^b [\xi f_1(x) - f_2(x)]^2\, dx = \xi^2 \int_a^b f_1^2(x)\, dx - 2\xi \int_a^b f_1(x) f_2(x)\, dx + \int_a^b f_2^2(x)\, dx,$$

where ξ is any real number. The form of the left-hand side implies that this cannot be negative for any real ξ. But if an expression of the form $A\xi^2 - 2\xi B + C$ is non-negative for all real ξ, we know from elementary algebra that $AC - B^2 \geqslant 0$. On applying this to the right-hand side above we get (126_2). The inequality was first proved for

integrals by V. L. Buniakowski. It was encountered by Cauchy for sums.

30. Properties of scalar products and norms. We now mention some properties of scalar products and norms. On applying (126_1) and taking into account that $|\bar{y}_k| = |y_k|$, we can write:

$$|(\mathbf{x}, \mathbf{y})|^2 = \left|\sum_{k=1}^{n} x_k \bar{y}_k\right|^2 \leqslant \sum_{k=1}^{n} |x_k|^2 \cdot \sum_{k=1}^{n} |y_k|^2,$$

i.e.

$$|(\mathbf{x}, \mathbf{y})| \leqslant \|\mathbf{x}\| \cdot \|\mathbf{y}\|. \tag{128}$$

We now prove the so-called triangle rule:

$$\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|. \tag{129}$$

We have:

$$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) = (\mathbf{x}, \mathbf{x}) + (\mathbf{y}, \mathbf{y}) + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}),$$

so that we get by taking into account (128):

$$\|\mathbf{x} + \mathbf{y}\|^2 \leqslant \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2,$$

whence (129) follows.

We conclude the present section by considering the effect of the choice of coordinate system on the metric of a space, i.e. on the expression for the square of the length of a vector. Suppose we choose a new system in place of the fundamental Cartesian system, with the fundamental set consisting of the independent vectors

$$\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \ldots \mathbf{z}^{(n)}.$$

We shall have for any vector:

$$\mathbf{x} = z_1 \mathbf{z}^{(1)} + \ldots + z_n \mathbf{z}^{(n)},$$

where the z_k are its components in the new coordinate system.

The square of the length of this vector will be given by its scalar product with itself, i.e.

$$|\mathbf{x}|^2 = \left(z_1 \mathbf{z}^{(1)} + \ldots + z_n \mathbf{z}^{(n)}, z_1 \mathbf{z}^{(1)} + \ldots + z_n \mathbf{z}^{(n)}\right).$$

On expanding in accordance with the formula previously given, we get the following expression for the square of the length of the vector:

$$|\mathbf{x}|^2 = \sum_{i, k=1}^{n} a_{ik} z_i z_k, \tag{130}$$

where the coefficients a_{ik} are given by

$$a_{ik} = \left(\mathbf{z}^{(i)}, \mathbf{z}^{(k)}\right).$$

These latter evidently take their conjugate values on interchanging the subscripts, i.e.

$$a_{ki} = \bar{a}_{ik}. \tag{131}$$

A sum of the form (130), with coefficients satisfying condition (131), is generally referred to as an *Hermitian form.* It follows at once that every expression of type (130) with condition (131) has real values only, whatever the complex z_k, since with $i \neq k$ a pair of terms of (130) will be conjugates, whilst in the case of terms of the form $a_{kk} \mid z_k \mid^2$, the a_{kk} are real by (131). Furthermore, we can say from the method of obtaining Hermitian form (130) that it is not negative and only vanishes when all the z_k vanish. Formula (130) in fact defines the metric of the space in the new coordinate system.

Metric (130) will coincide with metric (110) in the corresponding Cartesian system if

$$a_{ik} = 0 \quad \text{for} \quad i \neq k \quad \text{and} \quad a_{kk} = 1$$

or

$$\left(\mathbf{z}^{(i)}, \mathbf{z}^{(k)}\right) = 0 \quad \text{for} \quad i \neq k \quad \text{and} \quad \left(\mathbf{z}^{(k)}, \mathbf{z}^{(k)}\right) = 1,$$

or in other words, if the $\mathbf{z}^{(k)}$ that we have taken as the fundamental set are mutually orthogonal unit vectors (of unit length).

A further point is that, if (113) defines a unitary transformation of vector components, the corresponding transformation for passing from the original to the new fundamental set is given by the matrix

$$U^{(*)-1},$$

the contragradient of U. This array coincides with $\overline{U}$ in the present case by (123), whilst it coincides simply with U in the case of real orthogonal transformations.

31. Orthogonalization of vectors. Suppose we are given any m linearly independent vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$. A set of vectors

$$C_1 \mathbf{x}^{(1)} + \ldots + C_m \mathbf{x}^{(m)},$$

where the C_k are arbitrary constants, defines our total space if $m = n$ or an m-dimensional subspace R_m if $m < n$. We show that m mutually orthogonal unit vectors can always be constructed so as to form the same subspace R_m as the vectors $\mathbf{x}^{(k)}$. In other words, these new

orthogonal unit vectors $\mathbf{z}^{(k)}$ must be expressible linearly in terms of the $\mathbf{x}^{(k)}$, and conversely, the $\mathbf{x}^{(k)}$ must be expressible in terms of the $\mathbf{z}^{(k)}$. We can construct these vectors in accordance with the following scheme:

$$\left.\begin{array}{l} \mathbf{y}^{(1)} = \mathbf{x}^{(1)} \\ \mathbf{y}^{(2)} = \mathbf{x}^{(2)} - \left(\mathbf{x}^{(2)},\ \mathbf{z}^{(1)}\right),\mathbf{z}^{(1)} \\ \mathbf{y}^{(3)} = \mathbf{x}^{(3)} - \left(\mathbf{x}^{(3)},\ \mathbf{z}^{(1)}\right)\mathbf{z}^{(1)} - \left(\mathbf{x}^{(3)},\ \mathbf{z}^{(2)}\right)\mathbf{z}^{(2)} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \end{array}\right\} \tag{132}$$

where

$$\mathbf{z}^{(1)} = \frac{\mathbf{y}^{(1)}}{|\mathbf{y}^{(1)}|}; \quad \mathbf{z}^{(2)} = \frac{\mathbf{y}^{(2)}}{|\mathbf{y}^{(2)}|}; \quad \dots; \quad \mathbf{z}^{(m)} = \frac{\mathbf{y}^{(m)}}{|\mathbf{y}^{(m)}|}. \tag{133}$$

The vector $\mathbf{z}^{(1)}$ is found from $\mathbf{y}^{(1)}$ simply by dividing by the length of $\mathbf{y}^{(1)}$ so that the length of $\mathbf{z}^{(1)}$ is unity. Next $\mathbf{y}^{(2)}$ is constructed in accordance with the above formula, and its definition implies at once that it is orthogonal to $\mathbf{z}^{(1)}$:

$$\left(\mathbf{y}^{(2)},\ \mathbf{z}^{(1)}\right) = \left(\mathbf{x}^{(3)},\ \mathbf{z}^{(2)}\right) - \left(\mathbf{x}^{(2)},\ \mathbf{z}^{(1)}\right)\left(\mathbf{z}^{(1)},\ \mathbf{z}^{(1)}\right) = 0.$$

On dividing $\mathbf{y}^{(2)}$ by its own length, we get $\mathbf{z}^{(2)}$. Next we construct $\mathbf{y}^{(3)}$ in accordance with (132), and it follows directly from the definition that it is orthogonal to $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$.

For we have by the orthogonality of $\mathbf{z}^{(1)}$ and $\mathbf{z}^{(2)}$:

$$\left(\mathbf{y}^{(3)},\ \mathbf{z}^{(2)}\right) = \left(\mathbf{z}^{(3)},\ \mathbf{z}^{(2)}\right) - \left(\mathbf{x}^{(3)},\ \mathbf{z}^{(2)}\right)\left(\mathbf{z}^{(2)},\ \mathbf{z}^{(2)}\right) = 0.$$

Division of $\mathbf{y}^{(3)}$ by its own length gives us $\mathbf{z}^{(3)}$, and so on.

All the newly constructed vectors are given linearly in terms of the $\mathbf{x}^{(k)}$. It is easily seen that conversely the $\mathbf{x}^{(k)}$ are expressible linearly in terms of the $\mathbf{z}^{(k)}$. We can do this simply by solving successively the above equations with respect to $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, and so on.

We notice also that none of the new $\mathbf{y}^{(k)}$ can be zero. For if we obtained some zero $\mathbf{y}^{(k)}$ at some step in the working, since this is given in terms of the $\mathbf{x}^{(s)}$ by a linear expression in which the coefficient of $\mathbf{x}^{(k)}$ is unity, we should now have linear dependence of certain of the $\mathbf{x}^{(s)}$, which contradicts our hypothesis that these vectors are linearly independent.

We recall that if pairs of a non-zero vector set are orthogonal, the vectors must be linearly independent.

If $m = n$, the $\mathbf{z}^{(k)}$ yield a complete system of orthogonal unit vectors, forming a Cartesian system. But if $m < n$, a complete Cartesian system requires the addition to the $\mathbf{z}^{(k)}$ of a further $(n - m)$ vectors, these latter being orthogonal both to each other and to

the $\mathbf{z}^{(k)}$. These new unit vectors must thus form a subspace R'_{n-m} of $(n-m)$ dimensions, orthogonal to the subspace R_m [12]. The new required vectors $\mathbf{u}$ must satisfy the system of equations

$$\left(\mathbf{u}, \mathbf{x}^{(1)}\right) = 0, \ldots, \left(\mathbf{u}, \mathbf{x}^{(m)}\right) = 0.$$

We have here a system of m homogeneous equations with n unknowns, the rank of the system being m, since the vectors $\mathbf{x}^{(k)}$ are linearly independent [12]. The system has $(n-m)$ linearly independent solutions, i.e. we get $(n-m)$ linearly independent vectors. On applying the above process of orthogonalization to these and reducing their lengths to unity, we obtain our complete set of linearly independent unit vectors.

We notice one further point. The subspace R_m formed by the orthogonal unit vectors $\mathbf{z}^{(k)}$ can equally be formed by another system of orthogonal unit vectors. We see this simply by applying a unitary transformation to the system of $\mathbf{z}^{(k)}$. Thus the process of orthogonalization can be carried out in different ways, and the method indicated above is only one of the possibilities.

32. Transformation of a quadratic form to a sum of squares. We take a second order surface in space with its centre at the coordinate origin:

$$Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Fyz + G = 0.$$

New axes (x', y', z') can always be chosen such that only terms containing squares of the coordinates remain in the transformed equation, i.e. such that the transformed equation has the form

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 + G = 0.$$

The problem amounts to finding the orthogonal transformation relating variables (x', y', z') and (x, y, z), so that the set of second degree coordinate terms on the left-hand side of the equation reduces to a sum of squares. We formulate the analogous problem for real n-dimensional space. Let us have the real quadratic form in n variables:

$$\varphi\left(x_1, \ldots, x_n\right) = \sum_{i,\,k=1}^{n} a_{ik} x_i x_k, \tag{134}$$

where the a_{ik} are real coefficients satisfying the condition

$$a_{ki} = a_{ik}. \tag{135}$$

We can take in the previous example, $x = x_1$, $y = x_2$, $z = x_3$ and $a_{11} = A, a_{22} = B, a_{33} = C, a_{12} = a_{21} = D, a_{13} = a_{31} = E, a_{23} = a_{32} = F.$

The matrix composed of elements a_{ik} is known as the matrix of the quadratic form (134). This is a symmetric matrix, i.e. the same as its transpose.

Suppose we transform (134) to new variables x'_k instead of x_k, the transformation being written as

$$(x_1, \ldots, x_n) = B(x'_1, \ldots, x'_n), \tag{136}$$

where B is a matrix with elements b_{ik}. On substituting in (134) from (136), we get

$$\varphi = \sum_{i,\,k=1}^{n} a_{ik}(b_{i1}x'_1 + \ldots + b_{in}x'_n)(b_{k1}x'_1 + \ldots + b_{kn}x'_n).$$

Removal of the brackets gives us for the coefficient of $x'_p\,x'_q$ with $p \neq q$:

$$\sum_{i,\,k=1}^{n} a_{ik}(b_{ip}b_{kq} + b_{iq}b_{kp}).$$

It is easily seen by using (135) that half the last expression is simply equal to the sum:

$$\sum_{i=k}^{n} b_{ip} \sum_{k=1}^{n} a_{ik}c_{kq}.$$

Hence, on dealing similarly with the case $p = q$, we see that the quadratic form becomes in the new variables:

$$\varphi = \sum_{i,\,k=1}^{n} c_{ik}x'_i x'_k, \tag{137}$$

where

$$c_{ik} = c_{ki} = \sum_{t=1}^{n} b_{ti} \sum_{s=1}^{n} a_{ts}b_{sk}.$$

Summation over s gives $\{AB\}_{ik}$. If we take t as indicating the column and i as indicating the row in the factor b_{ti}, b_{ti} will be the element $\{B^{(*)}\}_{it}$ of the transpose, whence

$$c_{ik} = c_{ki} = \sum_{t=1}^{n} \{B^{(*)}\}_{it}\{AB\}_{tk},$$

i.e. the transformed form (137) has a matrix given in terms of the matrix A of the form in the original variables and the matrix B of transformation (136) by

$$C = B^{(*)}AB. \tag{138}$$

If transformation (136) is orthogonal, the transpose $B^{(*)}$ of the orthogonal matrix B is the same as the inverse B^{-1}, and we have in this case, instead of (138):

$$C = B^{-1} AB. \tag{139}$$

Our task of finding an orthogonal transformation (136) reducing the quadratic form (134) to a sum of squares is thus equivalent to the task of finding an orthogonal matrix B such that matrix (139) is simply a diagonal matrix $[\lambda_1, \ldots, \lambda_n]$, inasmuch as, if a form reduces to a sum of squares, its matrix is in fact a diagonal matrix whose elements λ_k are the coefficients of the $x_k'^2$. We must therefore have, as previously:

$$B^{-1} AB = [\lambda_1, \ldots, \lambda_n]$$

or

$$AB = B[\lambda_1, \ldots, \lambda_n]. \tag{140}$$

We remark that the A here is real and symmetric, and not an arbitrary matrix, whilst B must be orthogonal. We shall proceed precisely as in [27], when considering the general case. We re-write (140) as

$$\sum_{s=1}^{n} a_{is} b_{sk} = \lambda_k b_{ik}. \tag{141}$$

We thus have n equations for the elements of the kth column of B. On introducing the vector $\mathbf{x}^{(k)}$ with components $(b_{1k}, \ldots, b_{nk})$, we can write the last equation as

$$A\mathbf{x}^{(k)} = \lambda_k \mathbf{x}^{(k)}. \tag{142}$$

On taking all the terms of (141) to one side, we have a system of n homogeneous equations for $b_{1k}, \ldots, b_{nk}$:

$$\left.\begin{array}{l} (a_{11} - \lambda_k) b_{1k} + a_{12} b_{2k} + \cdots + a_{1n} b_{nk} = 0 \\ a_{21} b_{1k} + (a_{22} - \lambda_k) b_{2k} + \cdots + a_{2n} b_{nk} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1} b_{1k} + a_{n2} b_{2k} + \cdots + (a_{nn} - \lambda_k) b_{nk} = 0 \end{array}\right\}. \tag{143}$$

The determinant of this system must vanish, and we get an algebraic equation of degree n for the numbers λ_k:

$$\begin{vmatrix} a_{11} - \lambda, & a_{12}, & \ldots, & a_{1n} \\ a_{21}, & a_{22} - \lambda, & \ldots, & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}, & a_{n2}, & \ldots, & a_{nn} - \lambda \end{vmatrix} = 0. \tag{144}$$

As we know, this is the characteristic equation of matrix A.

We show first of all that all the roots of equation (144) are real for a real symmetric matrix A. We start by indicating a new way of writing a quadratic form. Let $\mathbf{x}$ be a vector with real or complex components $(x_1, \ldots, x_n)$ and A a matrix with any elements a_{ik}. We form the scalar product:

$$(A\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{n} \bar{x}_i(a_{i1} x_1 + \ldots + a_{in} x_n) .$$

It can evidently be written in the form:

$$(A\mathbf{x}, \mathbf{x}) = \sum_{i, k=1}^{n} a_{ik} \bar{x}_i x_k . \tag{145}$$

If the condition

$$a_{ki} = \bar{a}_{ik} \ (a_{kk} - \text{real}) , \tag{146}$$

is satisfied, we have an Hermitian form, the value of which is necessarily real. The case of real symmetric A is a particular case of condition (146). If in addition the components of $\mathbf{x}$ are real, (145) in fact yields the quadratic form (134).

We now turn to the proof that the roots of (144) are real. Let λ_k be a root of the equation. System (143) now gives us the the components of a vector $\mathbf{x}^{(k)}$ (real or complex) which satisfies equation (142). We take the scalar product on the right with $\mathbf{x}^{(k)}$ of both sides of this equation, and get:

$$|\mathbf{x}^{(k)}|^2 \lambda^k = (A\mathbf{x}^{(k)}, \mathbf{x}^{(k)}) .$$

The expression on the right is a real number, as we have seen, and consequently λ_k is also a real number. We have thus proved that the roots of (144) are real not only for a real symmetric matrix but also for any matrix whose elements satisfy condition (146). These latter matrices are generally described as *Hermitian*.

The coefficients of system (143) are real numbers in the present case, and we can take the components of $\mathbf{x}^{(k)}$ as real. We now show that if λ_p and λ_q are two different roots of (144), the corresponding $\mathbf{x}^{(p)}$ and $\mathbf{x}^{(q)}$ satisfying equation (142) are mutually orthogonal. We have by hypothesis:

$$A\mathbf{x}^{(p)} = \lambda_p \mathbf{x}^{(p)}; \quad A\mathbf{x}^{(q)} = \lambda_q \mathbf{x}^{(q)} .$$

On forming the scalar product of the first of these equations with $\mathbf{x}^{(q)}$, and of the second with $\mathbf{x}^{(p)}$, and subtracting, we get:

$$(A\mathbf{x}^{(p)}, \mathbf{x}^{(q)}) - (\mathbf{x}^{(p)}, A\mathbf{x}^{(q)}) = (\lambda_p - \lambda_q) \ (\mathbf{x}^{(p)}, \mathbf{x}^{(q)}). \tag{147}$$

We now show that, for any two (real or complex) vectors $\mathbf{x}$ and $\mathbf{y}$, we have

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y}) . \tag{148}$$

provided only that the elements of matrix A satisfy condition (146). The left-hand side of (148) gives, in fact:

$$(A\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{n} (a_{k1} x_1 + \ldots + a_{kn} x_n) \bar{y}_k = \sum_{i, k=1}^{n} a_{ki} x_i \bar{y}_k ,$$

or, by (146):

$$(A\mathbf{x}, \mathbf{y}) = \sum_{i, k=1}^{n} \bar{a}_{ik} x_i \bar{y}_k .$$

The right-hand side of (148) yields the same result. Our formula must likewise be valid for real orthogonal matrices, since these are a particular case of Hermitian matrices. By (148), the left-hand side of (147) is zero, and since $\lambda_p \neq \lambda_q$, we have $(\mathbf{x}^{(p)}, \mathbf{x}^{(q)}) = 0$, i.e. vectors $\mathbf{x}^{(p)}$ and $\mathbf{x}^{(q)}$ are in fact orthogonal. These vectors are real in the present case, and their orthogonality amounts to the condition that the sum of the products of their components is zero.

Thus if (144) has different roots, we have n mutually orthogonal real vectors $\mathbf{x}^{(k)}$. Equation (142) is linear and homogeneous in $\mathbf{x}^{(k)}$, and we can therefore multiply a solution of it by an arbitrary constant. It follows that we can take the lengths of the $\mathbf{x}^{(k)}$ as unity.

The components of these vectors form the columns of matrix B. In other words, B satisfies the condition for orthogonality with respect to columns, and is an orthogonal matrix. Hence our task of reducing a quadratic form to a sum of squares by means of an orthogonal transformation — or what amounts to the same thing, of reducing a matrix A to a diagonal form — has been solved, provided we assume that equation (144) has n different roots. The numbers λ_k are known as the *eigenvalues of matrix* A, whilst the $\mathbf{x}^{(k)}$ are the *eigenvectors of the matrix.*

33. The case of multiple roots of the characteristic equation. We now take the general case, when equation (144) can have multiple roots. We find the solution of (142) corresponding to a given root $\lambda = \lambda_1$ of (144). The solution will be a real vector $\mathbf{x}^{(1)}$ of unit length. We associate with it a further $(n - 1)$ real unit vectors, so that altogether a complete system of orthogonal unit vectors is formed [31]. The passage from the old to the new fundamental set will be expressed as usual by an orthogonal transformation of the vector components,

and A becomes the similar matrix $A_1 = B_1^{-1} A B_1$. The equation

$$A_1 \mathbf{x} = \lambda \mathbf{x} \tag{149}$$

corresponding to the new matrix, will have $\mathbf{x}^{(1)}$ as the solution corresponding to the eigenvalue $\lambda = \lambda_1$ (the eigenvalues are invariants in a similarity transformation), where $\mathbf{x}^{(1)}$ is the vector we took as the first of the fundamental set, so that its components are $(1, 0, \ldots, 0)$. On substituting this solution in (149), we get:

$$A_1(1, 0, \ldots, 0) = (\lambda_1, 0, \ldots, 0),$$

so that we have at once for the elements of the first column:

$$\{A_1\}_{11} = \lambda_1; \quad \{A_1\}_{21} = \{A_1\}_{31} = \ldots = \{A_1\}_{n1} = 0\,. \tag{150}$$

We now show that the real matrix A_1 is also symmetric, i.e. is the same as its transpose. In fact:

$$A_1^{(*)} = (B_1^{-1} A B_1)^{(*)} = B_1^{(*)} A^{(*)} B_1^{(*)-1}\,.$$

But since B_1 is orthogonal:

$$B_1^{(*)} = B_1^{-1} \text{ and } B_1^{(*)-1} = B_1\,,$$

whence it follows that

$$A_1^{(*)} = A_1\,.$$

On taking (150) into account, together with the symmetry of A_1, we can write:

$$\{A_1\}_{11} = \lambda_1; \quad \{A_1\}_{21} = \{A_1\}_{12} = \ldots = \{A_1\}_{n1} = \{A_1\}_{1n} = 0\,,$$

i.e. all the elements of the first row and first column of A_1 vanish with the exception, of course, of $\{A_1\}_{11} = \lambda_1$, i.e. A_1 has the form

$$A_1 = \begin{Vmatrix} \lambda_1, & 0, & \ldots, & 0 \\ 0, & a_{22}^{(1)}, & \ldots, & a_{2n}^{(1)} \\ \cdot & \cdot & \cdot & \cdot \\ 0, & a_{n2}^{(1)}, & \ldots, & a_{nn}^{(1)} \end{Vmatrix},$$

where we have written $a_{ik}^{(1)}$ for the elements of A_1.

The quadratic form φ becomes in the new variables:

$$\varphi = \lambda_1 y_1'^2 + \sum_{i,k=2}^{n} a_{ik}^{(1)} y_i' y_k'\,.$$

We have thus isolated one square and now have to consider the quadratic form in $(n-1)$ variables

$$\sum_{i,k=2}^{n} a_{ik}^{(1)} y_i' y_k'$$

or what is the same thing, we have to consider the $(n-1)$th order matrix C_1 corresponding to this and forming part of A_1. We can argue here exactly as above and find a unit vector $\mathbf{x}^{(2)}$ in the $(n-1)$ dimensional subspace formed by the remaining $(n-1)$ fundamental vectors such that $\mathbf{x}^{(2)}$ is a solution of the equation

$$C_1\,\mathbf{x}^{(2)} = \lambda_2\,\mathbf{x}^{(2)}$$

This vector is clearly orthogonal to $\mathbf{x}^{(1)}$. The fundamental vector $\mathbf{x}^{(1)}$ is preserved after this second transformation, whilst the remaining fundamental vectors become a new mutually orthogonal unit set, the first of which is $\mathbf{x}^{(2)}$. The quadratic form φ becomes in the new variables

$$\varphi = \lambda_1\,y_1''^2 + \lambda_2\,y_2''^2 + \sum_{i,\,k=3}^{n} a_{i,k}^{(2)}\,y_i''\,y_k''.$$

By proceeding thus, we finally reduce the quadratic form to a sum of squares, i.e. we reduce the corresponding matrix to a diagonal form. We do this as a result of applying a series of orthogonal transformations, and this is clearly equivalent to the application of the single orthogonal transformation B equal to their product.

The final diagonal matrix

$$B^{-1}AB = [\lambda_1, \ldots, \lambda_n] \tag{151}$$

is similar to the original matrix A, and consequently its characteristic equation

$$\begin{vmatrix} \lambda_1-\lambda, & 0, & \ldots, & 0 \\ 0, & \lambda_2-\lambda, & \ldots, & 0 \\ \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot & \cdot\;\cdot\;\cdot \\ 0, & 0, & \ldots, & \lambda_n-\lambda \end{vmatrix} = 0$$

is the same as (144), in other words, the coefficients λ_k of the squares in the reduced quadratic form

$$\varphi = \lambda_1\,x_1'^2 + \ldots + \lambda_n\,x_n'^2 \tag{152}$$

are the roots of (144), each multiple root being repeated as many times as its multiplicity.

As we know, each column of the final orthogonal transformation B yields a vector which is a solution of (142), where the method of derivation implies that the λ_k corresponding to a given column is the same as the coefficient of the corresponding variable in quadratic

form (152). The situation may be indicated more precisely. By (136), the orthogonal transformation B, which satisfies condition (140), transforms the variables $(x'_1, \ldots, x'_n)$ to $(x_1, \ldots, x_n)$.

The inverse B^{-1} is the transpose of B, i.e. we have

$$x'_k = b_{1k}x_1 + \ldots + b_{nk}x_n \quad (k = 1, 2, \ldots, n) \tag{153}$$

and $\mathbf{x}^{(k)}$, with components $(b_{1k}, \ldots, b_{nk})$, is a solution of (142) with $\lambda = \lambda_k$.

We finally show that we have found *all the solutions of* (142). First of all, it follows from the above arguments that λ_k must be a solution of (144). Let λ be any root of this equation; we suppose for clarity that it has multiplicity three, whilst it is obviously possible to take $\lambda = \lambda_1 = \lambda_2 = \lambda_3$. The equation

$$A\mathbf{x} = \lambda_1 \mathbf{x} \tag{154}$$

now has, by the previous discussion, the three solutions:

$$\mathbf{x}^{(1)}(b_{11}, \ldots, b_{n1}); \quad \mathbf{x}^{(2)}(b_{12}, \ldots, b_{n2}); \quad \mathbf{x}^{(3)}(b_{13}, \ldots, b_{n3}).$$

We show that every solution of (154) must be a linear combination of these. If this were not so, we should in fact have a solution $\mathbf{y}$ linearly independent of $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}$. Our $\mathbf{y}$ could be complex, but in this case its real and imaginary parts have to satisfy (154) individually, since the coefficients of the equation are real. At least one of these parts must obviously be a vector linearly independent of the $\mathbf{x}^{(k)}$ $(k = 1, 2, 3)$. Hence we can suppose that our $\mathbf{y}$ is real. As shown above, it must be orthogonal to all the $\mathbf{x}^{(k)}$ with $k > 3$, since the λ_k corresponding to these are different from λ_1. The result is that $\mathbf{y}$ is linearly independent of the total set of $\mathbf{x}^{(k)}$, i.e. we have $(n + 1)$ linearly independent vectors, which is impossible. Every m-tuple root of (144) thus implies precisely m linearly independent solutions of (154).

Substitution of an m-tuple root $\lambda = \lambda_0$ for λ_k in the coefficients of system (143) gives us a homogeneous system with m linearly independent solutions, i.e. the rank of the system must be $(n - m)$. In other words, the system reduces to $(n - m)$ equations. We take any solution of the system and multiply it by a factor such that the sum of the squares of the numbers appearing in the solution is unity. This gives us one vector corresponding to the root $\lambda = \lambda_0$. To find the next vector, we add to the $(n - m)$ equations of our system a

further equation expressing the orthogonality of the required vector to that already obtained. We thus have a homogeneous system of $(n - m + 1)$ equations for the components of the new vector. On taking any solution of this system and normalizing again (reducing the vector length to unity), we are faced with the task of finding a third vector corresponding to the root $\lambda = \lambda_0$. We do this by adding to the original $(n - m)$ equations a further two equations, expressing the orthogonality of the new required vector to the two already found, and so on until we arrive at the total set of m mutually orthogonal unit vectors corresponding to the m-tuple root $\lambda = \lambda_0$. A direct consequence of this method of construction is a certain arbitrariness in constructing the basic solutions of equations (142). If all the roots of the equation are simple, this arbitrariness merely amounts to the possibility of multiplying all the components of $\mathbf{x}^{(k)}$ by (-1). Now let (144) have an m-tuple root. In this case the corresponding m orthogonal unit vectors making up the solution of (142) form an m-dimensional subspace R_m. We can obviously make an arbitrary choice of mutually orthogonal fundamental vectors in this subspace, and they will likewise be solutions of (142) with $\lambda = \lambda_0$, i.e. we can pass from one set of orthogonal normalized solutions to another by carrying out an orthogonal transformation of R_m. All these remarks equally apply to any other multiple root of (144).

What has been said may be explained by returning to the problem treated at the start of the previous section, of reducing the equation of a second order surface to axes of symmetry. Suppose for definiteness that the surface is an ellipsoid. The case of different roots of (144) corresponds to the fact that all the semi-axes of the ellipsoid are different. In this case the natural arbitrariness in the choice of final coordinate axes amounts to a change in the direction of these axes. If (144), which is here an equation of the third degree, has two equal roots, the ellipsoid becomes an ellipsoid of revolution, and two axes of symmetry can lie where we please in the plane passing through the centre and perpendicular to the axis of revolution, provided only that they are perpendicular to each other, i.e. in the present case the arbitrariness in the choice of final axes consists further in an arbitrary orthogonal transformation in the above-mentioned plane. Finally, if all three roots of (144) are equal, our ellipsoid is a sphere, and our equation does not contain coordinate product terms. Here our choice of Cartesian axes in space is in general completely arbitrary.

34. Examples. We shall consider two numerical examples.

1. To reduce the surface given by

$$x_1^2 + 5x_2^2 + x_3^2 + 2x_1 x_2 + 6x_1 x_3 + 2x_2 x_3 = 5$$

to axes of symmetry.

The corresponding quadratic form will be

$$\begin{aligned}\varphi = x_1^2 + x_1 x_2 + 3x_1 x_3 + \\ + x_2 x_1 + 5x_2^2 + x_2 x_3 + \\ + 3x_3 x_1 + x_3 x_2 + x_3^2 .\end{aligned}$$

The characteristic equation of its matrix is

$$\begin{vmatrix} 1-\lambda, & 1, & 3 \\ 1, & 5-\lambda, & 1 \\ 3, & 1, & 1-\lambda \end{vmatrix} = 0 ,$$

whence, on expanding by the first row:

$$(1-\lambda)\,[(5-\lambda)(1-\lambda) - 1] - (1 - \lambda - 3) + 3\,[1 - 3\,(5-\lambda)] = 0$$

or

$$\lambda^3 - 7\lambda^2 + 36 = 0 .$$

It is easily verified that the roots of this equation are

$$\lambda_1 = -2; \quad \lambda_2 = 3; \quad \lambda_3 = 6 ,$$

and the equation of our surface, referred to axes of symmetry, is

$$-2x_1'^2 + 3x_2'^2 + 6x_3'^2 = 5 .$$

We shall now find the elements of the orthogonal matrix:

$$B = \begin{Vmatrix} b_{11}, & b_{12}, & b_{13} \\ b_{21}, & b_{22}, & b_{23} \\ b_{31}, & b_{32}, & b_{33} \end{Vmatrix} .$$

We have the system for these,

$$\left.\begin{aligned} (1-\lambda)\, b_{1k} + b_{2k} + 3b_{3k} = 0 \\ b_{1k} + (5-\lambda)\, b_{2k} + b_{3k} = 0 \\ 3b_{1k} + b_{2k} + (1-\lambda)\, b_{3k} = 0 . \end{aligned}\right\} \qquad (155)$$

We first substitute $\lambda = \lambda_3 = -2$, which leads us to two equations:

$$\begin{aligned} 3b_{11} + b_{21} + 3b_{31} = 0 \\ b_{11} + 7b_{21} + b_{31} = 0 . \end{aligned}$$

The solution of this system has the form

$$b_{11} = -k_1; \quad b_{21} = 0; \quad b_{31} = k_1 ,$$

where k_1 is an arbitrary number. We choose it so that the sum of the squares of the numbers making up the solution is equal to unity. We finally have

$$b_{11} = \frac{1}{\sqrt{2}} ; \quad b_{21} = 0; \quad b_{31} = -\frac{1}{\sqrt{2}} ,$$

where the solution can be taken with the opposite sign.

We now substitute $\lambda = \lambda_2 = 3$ in the coefficients of system (155). We get a system in which the third equation is the difference between the first two, so that it reduces to two equations:

$$\begin{aligned} -2b_{12} + b_{22} + 3b_{32} &= 0 \\ b_{12} + 2b_{22} + b_{32} &= 0 . \end{aligned}$$

The solution of this system, normalized to unity, is easily found:

$$b_{12} = \frac{1}{\sqrt{3}}; \quad b_{22} = -\frac{1}{\sqrt{3}}; \quad b_{32} = \frac{1}{\sqrt{3}} .$$

We finally substitute the third root in the coefficients of (155). Again we get a system in which one equation is a consequence of the other two. On solving the two remaining equations and normalizing to unity, we have:

$$b_{13} = \frac{1}{\sqrt{6}}; \quad b_{23} = \frac{2}{\sqrt{6}}; \quad b_{33} = \frac{1}{\sqrt{6}} .$$

The transformations of the variables are given in the present case by

$$\begin{aligned} x_1' &= \frac{1}{\sqrt{2}} x_1 \qquad\qquad\quad - \frac{1}{\sqrt{2}} x_3 \\ x_2' &= \frac{1}{\sqrt{3}} x_1 - \frac{1}{\sqrt{3}} x_2 + \frac{1}{\sqrt{3}} x_3 \\ x_3' &= \frac{1}{\sqrt{6}} x_1 + \frac{2}{\sqrt{6}} x_2 + \frac{1}{\sqrt{6}} x_3 . \end{aligned}$$

2. To reduce to axes of symmetry the surface

$$2x_1^2 + 6x_2^2 + 2x_3^2 + 8x_1 x_3 = 1 .$$

Here the quadratic form is written as

$$\begin{aligned} \varphi = 2x_1^2 + 0x_1 x_2 + 4x_1 x_3 &+ \\ + 0x_2 x_1 + 6x_2^2 \quad + 0x_2 x_3 &+ \\ + 4x_3 x_1 + 0x_3 x_2 + 2x_3^2 &, \end{aligned}$$

The characteristic equation of its matrix is

$$\begin{vmatrix} 2-\lambda, & 0, & 4 \\ 0, & 6-\lambda, & 0 \\ 4, & 0, & 2-\lambda \end{vmatrix} = 0 .$$

Expansion of the determinant gives us the equation

$$\lambda^3 - 10\lambda^2 + 12\lambda + 72 = 0 .$$

The roots are

$$\lambda_1 = -2; \quad \lambda_2 = \lambda_3 = 6,$$

i.e. the equation has a double root.

Next we find the coefficients of the orthogonal transformation, for which we have the system

$$\left.\begin{aligned} (2-\lambda)\, b_{1k} \quad + \qquad 4b_{3k} &= 0 \\ (6-\lambda)\, b_{2k} \qquad\quad &= 0 \\ 4b_{1k} \quad + (2-\lambda)\, b_{3k} &= 0. \end{aligned}\right\} \tag{155_1}$$

On substitutitng $\lambda = -2$, we easily arrive at the normalized solution

$$b_{11} = \frac{1}{\sqrt{2}}; \quad b_{21} = 0; \quad b_{31} = -\frac{1}{\sqrt{2}}.$$

We now substitute the double root $\lambda = 6$ in the coefficients of system (155_1), for which we must get two linearly independent and mutually orthogonal solutions. The substitution leads us to the single equation

$$-b_{12} + b_{32} = 0.$$

We take the normalized solution of this equation:

$$b_{12} = \frac{1}{\sqrt{2}}; \quad b_{22} = 0; \quad b_{32} = \frac{1}{\sqrt{2}}.$$

As regards the second solution, we notice that it has to satisfy both system (155_1) and the condition for orthogonality with the solution already obtained. This gives us two equations for it:

$$-b_{13} + \qquad b_{33} = 0$$

$$\frac{1}{\sqrt{2}} b_{13} + \frac{1}{\sqrt{2}} b_{33} = 0$$

or

$$b_{13} = b_{33} = 0,$$

whence the normalized solution becomes

$$b_{13} = 0; \quad b_{23} = 1; \quad b_{33} = 0.$$

Finally, the orthogonal transformation will be

$$\begin{aligned} x_1' &= \frac{1}{\sqrt{2}} x_1 - \frac{1}{\sqrt{2}} x_3 \\ x_2' &= \frac{1}{\sqrt{2}} x_1 + \frac{1}{\sqrt{2}} x_3 \\ x_3' &= x_2 \end{aligned}$$

and the surface has the equation, referred to axes of symmetry:

$$-2x_1'^2 + 6\,(x_2'^2 + x_3'^2) = 1.$$

35. Classification of quadratic forms. The problem of reducing a quadratic form to a sum of squares can be posed in a more general form to that given above, where we have required orthogonality of

the linear transformation from the new variables to the old. We can take the following more general problem: to reduce the real quadratic form (134) to the form

$$\varphi = \mu_1 X_1^2 + \mu_2 X_2^2 + \ldots + \mu_n X_n^2, \tag{156}$$

where the X_k are n linearly independent real linear forms in the variables x_k. The coefficients μ_k in this statement of the problem are not definite numbers such as we had above, though we can say something about them, viz., the number of non-zero μ_k must always be equal to the rank of the matrix composed of the coefficients a_{ik} of the quadratic form. In other words, in any reduction of a quadratic form to a sum of squares of linearly independent linear forms, the number of the squares is equal to the rank of the matrix just mentioned. In addition to this, a further property holds, which is usually known as *the law of inertia of quadratic forms:* in any transformation of a real quadratic form to the form (156), where the linear forms X_k are also real, *the number of positive coefficients μ_k (and the number of negative coefficients μ_k) is always the same.* We shall prove these assertions at the end of the present section.

This general problem of reducing a quadratic form to form (156) is always very easily solved on separating out perfect squares. We shall do this in the particular example:

$$\varphi = x_1^2 + 4x_2^2 + x_3^2 + 2x_1x_2 - 6x_1x_3 + 8x_2x_3.$$

We obtain a perfect square from $(x_1^2 + 2x_1x_2 - 6x_1x_3)$ by adding $(x_2^2 + 9x_3^2 - 6x_2x_3)$, when we can write φ in the form

$$\varphi = (x_1 + x_2 - 3x_3)^2 + 3x_2^2 - 8x_3^2 + 14x_2x_3.$$

We separate out a further perfect square in the same way, and can finally write our quadratic form in form (156):

$$\varphi = (x_1 + x_2 - 3x_3)^2 - 2\left(2x_3 - \frac{7}{4}x_2\right)^2 + \frac{73}{8}(x_2)^2.$$

The linear forms appearing in the brackets are clearly linearly independent.

The working is somewhat different if squares of the variables are absent in the expression for φ. Suppose we have

$$\varphi = ax_1x_2 + Px_1 + Qx_2 + R,$$

where a is a numerical coefficient differing from zero, P and Q are linear forms of variables, not including x_1 and x_2, and R is a quadratic form which likewise does not include x_1 and x_2. We can write

$$\varphi = a\left(x_1 + \frac{Q}{a}\right)\left(x_2 + \frac{P}{a}\right) + R - \frac{PQ}{a}.$$

If we set

$$X_1 = \frac{1}{2}\left(x_1 + x_2 + \frac{P+Q}{a}\right); \quad X_2 = \frac{1}{2}\left(x_1 - x_2 - \frac{P-Q}{a}\right)$$

and

$$\varphi_1 = R - \frac{PQ}{a},$$

we get

$$\varphi = aX_1^2 - aX_2^2 + \varphi_1,$$

where φ_1 is a quadratic form which does not include x_1 and x_2. By separating out these two squares, we have got rid of two variables.

The reduction of a quadratic form to form (156) makes it possible to classify the form as follows:

I. Let all the coefficients μ_k in (156) be positive. In this case the form is said to be *positive definite*. It may easily be seen to have positive values for all real x_k and to vanish only when all the x_k vanish. For, since all the μ_k are positive, the necessary and sufficient condition for the vanishing of the right-hand side of (156) is that all the linear forms x_k vanish. We thus get a system of n homogeneous equations for the x_k with a non-zero determinant (the forms are linearly independent), so that only the zero solution exists.

II. If all the μ_k are negative, the quadratic form is said to be *negative definite*. As above, it can be seen to have only negative values for all real x_k and to vanish only when all the x_k vanish.

III. We now take the case when some of the μ_k vanish, though all the remainder have the same sign, say positive. Our form φ now becomes

$$\varphi = \mu_1 X_1^2 + \ldots + \mu_m X_m^2 \qquad (m < n), \tag{156$_1$}$$

where all the μ_k are positive. Here again the form cannot be negative whatever the values of the x_k, though it can vanish for non-zero x_k. For if we want to find the zeros of the form, we have to write a system of m homogeneous equations in x_k:

$$X_1 = X_2 = \ldots = X_m = 0,$$

and since $m < n$, this system certainly has non-zero solutions. Similarly, if all the μ_k are negative in (156_1), the quadratic form cannot take positive values, though it can vanish for non-zero x_k. Here the form is said to be positive or negative semi-definite.

IV. Finally, if we get both positive and negative coefficients μ_k in (156), the form may easily be shown to take both positive and negative values for real x_k. It is described as *alternating* in this case.

The above classification of real quadratic forms has an immediate application to the problem of the maxima and minima of functions of several variables. Let us take the function of n independent variables $x_1, \ldots, x_n$:

$$\psi(x_1 \ldots, x_n),$$

and let the necessary conditions for maxima and minima be satisfied for $x_1 = \ldots = x_n = 0$, i.e. all the partial derivatives of ψ with respect to the independent variables vanish at the origin. We have on expanding our function in a Maclaurin series:

$$\psi(x_1, \ldots, x_n) - \psi(0, \ldots, 0) = \varphi(x_1, \ldots, x_n) + \omega,$$

where we have written $\varphi(x_1, \ldots, x_n)$ for the quadratic form in the variables x_k, and ω for the set of terms of higher order than the second in the x_k. If the quadratic form φ is positive definite, we have a minimum of the function at $x_1 = \ldots = x_n = 0$. If it is negative definite, we have a maximum. If it is alternating, we neither have a maximum nor a minimum, and finally, if φ is positive or negative semi-definite, we have a doubtful case. This result is the natural complement of that obtained in [I, 133] for functions of two independent variables.

We turn to the proof of the statements made at the beginning of the present section. Let us take the quadratic form

$$\varphi = \sum_{i,k=1}^{n} a_{ik}x_i x_k \qquad (a_{ik} = a_{ki}),$$

the rank of the matrix of its coefficients being r. We compose the system of n linear forms:

$$\frac{1}{2}\frac{\partial\varphi}{\partial x_s} = \sum_{l=1}^{n} a_{sl}\, x_l \ (s = 1, 2, \ldots, n). \tag{157}$$

We have used the conditions $a_{ik} = a_{ki}$ in forming the expressions for these partial derivatives. Obviously r is the rank of system (157) in the sense described in [11].

Suppose that φ is reduced to the sum of the squares of m linearly independent forms y_s:

$$y_s = \beta_{s1}x_1 + \beta_{s2}x_2 + \cdots + \beta_{sn}x_n. \tag{158}$$

i.e.

$$\varphi = \mu_1 y_1^2 + \mu_2 y_2^2 + \cdots + \mu_m y_m^2, \tag{159}$$

where the μ_s differ from zero. We have to show that $m = r$. We use (159) to obtain the linear forms (157):

$$\frac{1}{2}\frac{\partial\varphi}{\partial x_s} = \mu_1\beta_{1s}y_1 + \mu_2\beta_{2s}y_2 + \cdots + \mu_m\beta_{ms}y_m \quad (s = 1, 2, \ldots, n). \tag{157_1}$$

The variables y_s can take any values since forms (158) are linearly independent. Hence by definition of the linear dependence of forms (157_1) the y_s can be taken as independent variables, and the greatest number of linearly independent forms in system (157_1) must be equal to the rank of the matrix of coefficients $\mu_k \beta_{ki}$, where the column subscript k takes all the values $k = 1, 2, \ldots, m$, and the row subscript i all the values $i = 1, 2, \ldots, n$. The elements of each column of the matrix have the common factor μ_k which is non-zero, and hence the rank of the matrix of $\mu_k \beta_{ki}$ is the same as the rank of the matrix of β_{ki}. Since (158) is a system of m linearly independent forms, this rank is m, i.e. the greatest number of linearly independent forms in system (157_1) or (157) is m. On the other hand, this number is r by hypothesis, whence it follows that $m = r$.

We now show that the number of positive (and negative) coefficients μ_s is always the same, whatever the method of writing φ as an expression of the type (159), where the y_s are real linearly independent forms. We shall assume the opposite and prove a contradiction. Thus let φ be expressed by two formulae of type (159) in which the number of positive coefficients is not the same:

$$\left.\begin{aligned} \varphi &= \lambda_1 y_1^2 + \ldots + \lambda_p y_p^2 - \lambda_{p+1} y_{p+1}^2 - \ldots - \lambda_m y_m^2, \\ \varphi &= \lambda_1' y_1'^2 + \ldots + \lambda_p' y_p'^2 - \lambda_{q+1}' y_{q+1}'^2 - \ldots - \lambda_m' y_m^2. \end{aligned}\right\} \tag{160}$$

The λ_s' and λ_s in these expressions are assumed positive. The forms $y_1, \ldots, y_m$ are linearly independent, and the same can be said of $y_1', \ldots, y_m'$. Since $p \neq v$, we can always take say $p < q$. We show that this leads to an absurdity. We associate the forms $y_{m+1}, \ldots, y_n$ with $y_1, \ldots, y_m$ so as to obtain a complete system of linearly independent forms [11]. We write down the system of linear homogeneous equations for $x_1, x_2, \ldots, x_n$:

$$y_1 = 0; \ldots ; \quad y_p = 0; \quad y_{p+1}' = 0; \ldots ; \quad y_m' = 0; \quad y_{m+1} = 0; \ldots ; \quad y_n = 0. \tag{161}$$

The number of these homogeneous equations is

$$p + (m - q) + (n - m) = n - (q - p),$$

and, since $q - p > 0$, this number is less than n. Consequently the system written has real non-zero solutions. We take any one of the solutions: $x_s = x_s^{(0)}$ $(s = 1, 2, \ldots, n)$. With these values of x_s, we have by (161):

$$\varphi = -\lambda_{p+1} y_{p+1}^2 - \ldots - \lambda_m y_m^2 = \lambda_1' y_1'^2 + \ldots + \lambda_q' y_q'^2.$$

It is clear from this that φ must vanish for $x_s = x_s^{(0)}$, and these x_s must therefore satisfy, in addition to equations (161):

$$y_{p+1} = 0; \ \ldots ; \ y_m = 0.$$

We see finally that the complete system of linearly independent forms, $y_1, \ldots, y_n$, must vanish for $x_s = x_s^{(0)}$. But this is impossible, inasmuch as the linear independence of the forms y_s implies the non-vanishing of the determinant of the system

$$y_1 = 0; \quad y_2 = 0; \ \ldots ; \ y_n = 0,$$

which is homogeneous in $x_1, x_2, \ldots, x_n$. We have proved the law of inertia by arriving at this contradiction.

36. Jacobi's formula. Jacobi's formula, which offers a convenient form of the reduction of a quadratic form to a sum of squares, will be stated without proof.

We first introduce the following notation:

$$A_i(x) = \sum_{k=1}^{n} a_{ik} x_k \qquad (i = 1, 2, \ldots, n),$$

$$\varDelta_0 = 1; \quad \varDelta_1 = a_{11}; \quad \varDelta_k = \begin{vmatrix} a_{11}, & a_{12}, & \ldots, & a_{1k} \\ a_{21}, & a_{22}, & \ldots, & a_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & a_{k1}, & \ldots, & a_{kk} \end{vmatrix},$$

$$(k = 2, 3, \ldots, n)$$

$$X_1 = A_1(x); \; X_k = \begin{vmatrix} a_{11}, & \ldots, & a_{1k-1}, & A_1(x) \\ a_{21}, & \ldots, & a_{2k-1}, & A_2(x) \\ \cdot & \cdot & \cdot & \cdot \\ a_{k1}, & \ldots, & a_{kk-1}, & A_k(x) \end{vmatrix}.$$

If the rank of the matrix of a_{ik} is r, and the determinants $\varDelta_1, \varDelta_2, \ldots, \varDelta_r$ do not vanish, Jacobi's formula becomes:

$$\varphi = \sum_{i,k=1}^{n} a_{ik} x_i x_k = \sum_{k=1}^{r} \frac{X_k^2}{\varDelta_k \varDelta_{k-1}}, \tag{162}$$

where the linear forms X_k $(k = 1, 2, \ldots, r)$ are linearly independent. The formula makes it possible to see from the signs of the $\varDelta_k$ to what class φ belongs as regards the law of inertia.

In particular, if all the determinants $\varDelta_1, \varDelta_2, \ldots, \varDelta_n$ are positive (with this, $r = n$), it follows from (162) that φ is positive definite. The converse can be shown: if φ is a positive definite form, all the determinants must be positive. The variables x_s can be enumerated in any order, of course, when applying (162). The $\varDelta_k$ naturally also change on changing the enumeration, and each of the principal minors of the matrix $|\, a_{ik} \,|_1^n$ can be a determinant of the sequence $\varDelta_k$ for a given enumeration of the x_s. It follows from what has been said above that all the principal minors are positive in the case of a positive definite form φ, but it is sufficient here to verify the positiveness of the determinants

$$\varDelta_s \; (s = 1, 2, \ldots, n).$$

It can be shown that the necessary and sufficient condition for φ to be positive (of constant sign) is for all the principal minors to be non-negative, i.e. they can be greater than or equal to zero. It is not sufficient in this case to find the signs of the determinants $\varDelta_s$ only, and the signs of all the principal minors have to be determined.

A proof of the statements of this section may be found in *Ostsilyatsionnie matritsi i malie kolebaniya mechanicheskich system* ("Oscillation matrices and the small vibrations of mechanical systems") by F. R. Gantmaker and M. G. Krein (1941).

37. The simultaneous reduction of two quadratic forms to sums of squares. Suppose we have the two quadratic forms

$$\varphi_1 = \sum_{i,\,k=1}^{n} a_{ik}x_ix_k; \qquad \varphi_2 = \sum_{i,\,k=1}^{n} b_{ik}x_ix_k,$$

where φ_1 is positive definite, i.e. reduces to the sum of n positive squares. We require to find a linear transformation (not necessarily orthogonal) such that both forms are reduced to sums of squares.

We first of all introduce new variables y_k such that φ_1 reduces to a sum of squares. This can be done say by the elementary method indicated in the previous section. Our forms will become in the new variables:

$$\varphi_1 = \sum_{k=1}^{n} \mu_k y_k^2; \quad \varphi_2 = \sum_{i,\,k=1}^{n} b'_{ik}y_iy_k.$$

All the μ_k are positive by hypothesis, and we can bring in new real variables $z_k = \sqrt{\mu_k}\, y_k$. Now we have:

$$\varphi_1 = \sum_{k=1}^{n} z_k^2; \qquad \varphi_2 = \sum_{i,\,k=1}^{n} b''_{ik}z_iz_k.$$

We carry out an orthogonal transformation of the z_k to new variables z'_k, such that φ_2 reduces to a sum of squares.

Since the transformation is orthogonal here, φ_1 remains a sum of squares, and we have finally reduced both forms to sums of squares:

$$\varphi_1 = \sum_{k=1}^{n} z_k'^2; \quad \varphi_2 = \sum_{k=1}^{n} \lambda_k z_k'^2 .$$

The λ_k are sometimes called *the characteristic roots of form φ_2 with respect to form φ_1.*

We now establish the equation that has to be satisfied by these λ_k, and which is completely analogous to equation (144) of [32]. For this, we introduce the *discriminant* of a quadratic form, defined as follows: *the discriminant of a quadratic form is the determinant made up of its coefficients.*

Suppose we transform φ, the matrix of the coefficients of which is A, to new variables with the aid of the transformation

$$(x_1, \ldots, x_n) = B\,(x'_1, \ldots, x'_n).$$

As we know from [32], the matrix of the new form is

$$C = B^{(*)}AB,$$

and its determinant is given by

$$D(C) = D(B^{(*)})\,D(A)\,D(B).$$

The determinants $D(B^{(*)})$ and $D(B)$ are clearly equal since the corresponding matrices are obtained from each other by interchanging rows with columns. We thus have

$$D(C) = D(A)\,D(B)^2,$$

i.e. on linearly transforming the variables in a quadratic form its discriminant is multiplied by the square of the determinant of the linear transformation.

We now return to our quadratic forms φ_1, φ_2 and consider the form

$$\omega = \varphi_2 - \lambda\varphi_1 = \sum_{i,\,k=1}^{n} (b_{ik} - \lambda a_{ik})\,x_i x_k,$$

the coefficients of which contain the parameter λ.

After transformation to the new variables, this form becomes

$$\omega = \sum_{k=1}^{n} (\lambda_k - \lambda)\,z_k'^2,$$

and its discriminant in the new variables is evidently given by the product

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda)\ldots(\lambda_n - \lambda), \tag{163}$$

whilst the discriminant in the old variables is equal to the determinant with elements $(b_{ik} - \lambda\,a_{ik})$. As we have shown, these two discriminants differ only by a factor, viz., the square of the determinant of the transformation, which neither vanishes nor contains λ. It follows at once from this that both discriminants have the same roots with

respect to the parameter λ. On taking into account (162_1), we see that the numbers λ_k are the roots of the equation

$$\begin{vmatrix} b_{11}-\lambda a_{11}, b_{12}-\lambda a_{12}, \ldots, b_{1n}-\lambda a_{1n} \\ b_{21}-\lambda a_{21}, b_{22}-\lambda a_{22}, \ldots, b_{2n}-\lambda a_{2n} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ b_{n1}-\lambda a_{n1}, b_{n2}-\lambda a_{n2}, \ldots, b_{nn}-\lambda a_{nn} \end{vmatrix} = 0. \qquad (163_1)$$

38. Small vibrations. We saw above [II, 19] that the motion of a mechanical system with n degrees of freedom, the constraints of which do not contain time and which finds itself under the action of conservative forces, is given by the system of differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial q'_k}\right) - \frac{\partial T}{\partial q_k} = \frac{\partial U}{\partial q_k} \quad (k = 1, 2, \ldots, n). \qquad (164)$$

where T is the kinetic energy of the system and U the given force function of the q_k, which we take to be independent of t. As was mentioned previously, T is a quadratic form of the derivatives q'_k of q_k with respect to time:

$$T = \sum_{i,k=1}^{n} a_{ik}\, q'_i\, q'_k \qquad (a_{ki} = a_{ik}), \qquad (165)$$

the coefficients being given functions of the q_k. Suppose we have

$$\frac{\partial U}{\partial q_k} = 0 \quad \text{for} \quad q_1 = \ldots = q_n = 0 \qquad (k = 1, 2, \ldots, n). \qquad (166)$$

i. e. the partial derivatives of U vanish for $q_k = 0$.

With this, system (164) has the obvious solution $q_1 = \ldots = q_n = 0$, corresponding to a position of equilibrium. The function U is defined except for a constant, and we can always suppose that it vanishes for $q_1 = \ldots = q_n = 0$. We can therefore say, in view of (166), that the expansion of U in powers of q_k starts with a second order term. Let the quadratic form obtained from these second order terms be negative definite, whence it follows that U has a maximum for $q_1 = \ldots = q_n = 0$, or what amounts to the same thing, the potential energy $(-U)$ has a minimum. We proved in [II, 19] that the equilibrium position $q_1 = \ldots = q_n = 0$ is stable in this case, and for small initial excitations the system performs small vibrations about the equilibrium position, so that the q_k remain small throughout the motion. We can, therefore, assume when investigating these small vibrations, that U reduces simply to second order terms, i.e. has the form

$$-U = \sum_{i,k=1}^{n} b_{ik}\, q_i\, q_k \qquad (b_{ki} = b_{ik}). \qquad (167)$$

Similarly, in the coefficients a_{ik} of (165), we can take $q_k = 0$ approximately, so that the a_{ik} are specific numbers. On applying all this to system (164), we get a system of n linear equations with constant coefficients:

$$\left.\begin{aligned} &a_{11} q_1'' + a_{12} q_2'' + \cdots + a_{1n} q_n'' + b_{11} q_1 + b_{12} q_2 + \cdots + b_{1n} q_n = 0 \\ &a_{21} q_1'' + a_{22} q_2'' + \cdots + a_{2n} q_n'' + b_{21} q_1 + b_{22} q_2 + \cdots + b_{2n} q_n = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{n1} q_1'' + a_{n2} q_2'' + \cdots + a_{nn} q_n'' + b_{n1} q_1 + b_{n2} q_2 + \cdots + b_{nn} q_n = 0. \end{aligned}\right\} \tag{168}$$

If we seek a solution of this system in the form of harmonic vibrations of the same frequency and initial phase but with different amplitudes:

$$q_k = A_k \cos(\lambda t + \varphi) \qquad (k = 1, 2, \ldots, n), \tag{169}$$

substitution in (168) gives us a system of equations for the A_k and λ:

$$\left.\begin{aligned} &(b_{11} - \lambda^2 a_{11}) A_1 + (b_{12} - \lambda^2 a_{12}) A_2 + \cdots + (b_{1n} - \lambda^2 a_{1n}) A_n = 0 \\ &(b_{21} - \lambda^2 a_{21}) A_1 + (b_{22} - \lambda^2 a_{22}) A_2 + \cdots + (b_{2n} - \lambda^2 a_{2n}) A_n = 0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &(b_{n1} - \lambda^2 a_{n1}) A_1 + (b_{n2} - \lambda^2 a_{n2}) A_2 + \cdots + (b_{nn} - \lambda^2 a_{nn}) A_n = 0. \end{aligned}\right\} \tag{170}$$

The existence of a non-zero solution for the A_k requires the vanishing of the determinant of this system:

$$\begin{vmatrix} b_{11} - \lambda^2 a_{11}, & b_{12} - \lambda^2 a_{12}, & \ldots, & b_{1n} - \lambda^2 a_{1n} \\ b_{21} - \lambda^2 a_{21}, & b_{22} - \lambda^2 a_{22}, & \ldots, & b_{2n} - \lambda_2 a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} - \lambda^2 a_{n1}, & b_{n2} - \lambda^2 a_{n2}, & \ldots, & b_{nn} - \lambda^2 a_{nn} \end{vmatrix} = 0. \tag{171}$$

On taking a root of this equation and substituting in the coefficients of system (170), we get one or more solutions for the A_k, which we can then multiply by arbitrary constants. Moreover, (169) contains the arbitrary constant φ.

We get a clearer solution of the problem by applying the theory of quadratic forms. We first notice that quadratic form (165) in the variables q_k' is positive definite by its nature, inasmuch as it gives the kinetic energy of the motion. The problem furthermore implies that form (167) is positive definite. As we have seen, we can bring in new variables p_k by means of a linear transformation with constant coefficients of the old variables q_k such that both the forms T and $(-U)$ becomes sums of squares, where the coefficients of the squares in the case of T must be unity. We notice here that a linear relationship for the p_k and q_k leads to the same relationship between p_k' and q_k'. We thus have

$$T = \sum_{s=1}^{n} p_s'^2; \quad -U = \sum_{s=1}^{n} \lambda_s^2 p_s^2, \tag{172}$$

where all the coefficients of the p_s^2 are positive, so that we can write them as squares. Along with (168), we can write the Lagrange equation (164) for the new variables:

$$\frac{d}{dt}\left[\frac{\partial T}{\partial p_k'}\right] = \frac{\partial U}{\partial p_k}.$$

On substituting from (172), we get the extremely simple system:

$$p_k'' + \lambda_k^2 p_k = 0 \qquad (k = 1, 2, \ldots, n).$$

The solutions of this system are

$$p_k = C_k \cos(\lambda_k t + \varphi_k) \qquad (k = 1, 2, \ldots, n), \tag{173}$$

where C_k and φ_k are arbitrary constants. The generalized coordinates p_k are called the principal coordinates of the mechanical system.

The original coordinates q_k are given in terms of these by a linear transformation with constant coefficients. It follows at once from the results of the previous section that the λ_k must be the roots of equation (170). We remark that some of the roots may be equal, though even in this case (169) still gives the general solution of the problem of small vibrations within the context considered.

39. Extremal properties of the eigenvalues of quadratic forms. We consider the reduction of a quadratic form to a sum of squares from a new point of view. We confine ourselves to the case of three variables for the sake of simplicity:

$$\varphi = \sum_{i,k=1}^{3} a_{ik} x_i x_k = \sum_{k=1}^{3} \lambda_k x_k'^2, \tag{174}$$

where the x_k' and x_k are related by an orthogonal transformation:

$$\left.\begin{aligned} x_1 &= b_{11} x_1' + b_{12} x_2' + b_{13} x_3' \\ x_2 &= b_{21} x_1' + b_{22} x_2' + b_{23} x_3' \\ x_3 &= b_{31} x_1' + b_{32} x_2' + b_{33} x_3'. \end{aligned}\right\} \tag{175}$$

We shall suppose for definiteness that the λ_k are decreasing, i.e.

$$\lambda_1 > \lambda_2 > \lambda_3. \tag{176}$$

Our problem consists in determining the numbers λ_k and coefficients b_{ik} for values of the form φ on the unit sphere K, i.e. on the sphere with centre at the coordinate origin and unit radius:

$$x_1^2 + x_2^2 + x_3^2 = 1 \quad \text{or} \quad x_1'^2 + x_2'^2 + x_3'^2 = 1. \tag{177}$$

Each point of the sphere characterizes a certain direction in space, defined by the unit vector drawn from the origin to the point. We can write (174) as

$$\varphi = \lambda_1 (x_1'^2 + x_2'^2 + x_3'^2) + (\lambda_2 - \lambda_1) x_2'^2 + (\lambda_3 - \lambda_1) x_3'^2,$$

whence it is clear that we have on the unit sphere K:

$$\varphi = \lambda_1 + (\lambda_2 - \lambda_1) x_2'^2 + (\lambda_3 - \lambda_1) x_3'^2.$$

It follows at once from this that λ_1 is the maximum of φ on K.

The maximum is obviously obtained at the point

$$x_1' = 1; \quad x_2' = x_3' = 0,$$

or, by (175), at the point of K with the old coordinates

$$x_1 = b_{11}; \quad x_2 = b_{21}; \quad x_3 = b_{31}.$$

This point defines the vector corresponding to the first column of orthogonal transformation (175), i.e. the vector is the solution of the equation

$$A\mathbf{x} = \lambda\mathbf{x} \tag{178}$$

with $\lambda = \lambda_1$. Thus the eigenvalue of first magnitude of quadratic form (174) is equal to the maximum of the form on the unit sphere, whilst the corresponding eigenvector $\mathbf{x}^{(1)}$ is the solution of (178) which runs from the origin to the point of the unit sphere where the maximum occurs.

We now turn to finding the second eigenvalue and corresponding eigenvector. Let $x_1' = 0$ in the formulae. In this case we have the equation of a plane passing through the origin and perpendicular to the vector $\mathbf{x}^{(1)}$. The intersection of this plane with the unit sphere is the circle

$$x_2'^2 + x_3'^2 = 1.$$

We have on the circle:

$$\varphi = \lambda_2 x_2'^2 + \lambda_3 x_3'^2,$$

whence it is immediately clear that λ_2 is the maximum of φ on the unit sphere on condition that the corresponding vector is perpendicular to the $\mathbf{x}^{(1)}$ already found. We can show in the same manner as above that the corresponding vector $\mathbf{x}^{(2)}$, i.e. the solution of (178) for $\lambda = \lambda_2$, is the vector drawn to the point at which the maximum occurs.

Having obtained the two vectors, the third, $\mathbf{x}^{(3)}$, follows from the fact that it is perpendicular to both, whilst the eigenvalue λ_3 is the value of the form φ at the point of the unit sphere where it intersects $\mathbf{x}^{(3)}$.

If we had say $\lambda_1 = \lambda_2$, our search for the first maximum of φ on the unit sphere would lead us to an entire circle where the maximum is obtained, instead of a point.

The above discussion is easily carried over to the case of any number of dimensions. We shall merely state the result, which is completely analogous to the above. Suppose we have the real quadratic form in n variables:

$$\varphi = \sum_{i,\,k=1}^{n} a_{ik} x_i x_k. \tag{179}$$

A unit vector in real n-dimensional space is given by a set of real numbers, the sum of the squares of which is equal to unity. We shall say that the ends of these vectors lie on the unit sphere, the equation of which is obviously

$$x_1^2 + x_2^2 + \dots + x_n^2 = 1. \tag{180}$$

The highest characteristic root of the form φ will be the maximum of φ on the unit sphere (180), and the corresponding eigenvector is $\mathbf{x}^{(1)}$, drawn from the origin to the point where the maximum occurs. To get the next lower characteristic root, we consider the unit vectors perpendicular to the $\mathbf{x}^{(1)}$ already found. There will be an $\mathbf{x}^{(2)}$ among these, yielding the greatest value of φ. This second maximum λ_2 is equal to the second eigenvalue of the form, whilst $\mathbf{x}^{(2)}$ is the corresponding eigenvector. We now consider the unit vectors

perpendicular to $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, which is equivalent to associating with condition (180) the two further conditions:

$$(\mathbf{x}^{(1)}, \mathbf{x}) = 0 \quad \text{and} \quad (\mathbf{x}^{(2)}, \mathbf{x}) = 0.$$

One vector among these again yields a greatest value of φ, this being the eigenvalue λ_3 that comes third in magnitude, the vector in question being the corresponding eigenvector, and so on.

We could have arranged the eigenvalues of the quadratic form in increasing, instead of decreasing, order, so that the first would be the least, the next the second higher, and so on. This would lead to a precisely analogous problem to the above, except that a reference to least value would have to be substituted for every reference to greatest value.

All the above arguments may likewise be generalized to the case of simultaneous reduction of two quadratic forms to the sums of squares. Let the two quadratic forms

$$\varphi = \sum_{i,k=1}^{n} a_{ik} x_i x_k; \quad \psi = \sum_{i,k=1}^{n} b_{ik} x_i x_k$$

reduce to the sums of squares:

$$\varphi = \sum_{k=1}^{n} x_k'^2; \qquad \psi = \sum_{k=1}^{n} \lambda_k x_k'^2$$

with the aid of the linear transformation

$$(x_1, \ldots, x_n) = B\,(x_1', \ldots, x_n'),$$

the λ_k above being assumed to occur in decreasing order.

With this, λ_1 is the greatest value of ψ on condition that

$$\varphi = 1,$$

this greatest value being in fact obtained for

$$x_1 = b_{11}; \quad x_2 = b_{21}; \ldots; \quad x_n = b_{n1}.$$

The succeeding eigenvalues may be similarly determined.

40. Hermitian matrices and Hermitian forms. We have considered real symmetric matrices in the above sections and have noted that they represent a particular case of Hermitian matrices in which the elements are complex numbers satisfying

$$a_{ki} = \bar{a}_{ik}. \tag{181}$$

Setting $i = k$, this relationship shows that the diagonal elements a_{kk} must be real.

An alternative definition of Hermitian matrix is as follows: an Hermitian matrix is unchanged if its rows and columns are inter-

changed and all its elements are replaced by their conjugates, i.e. in the notation of [26]:

$$\overline{A}^{(*)} = A \text{ or } \quad \tilde{A} = A. \tag{182}$$

As already mentioned, $\tilde{A}$ is called the Hermitian conjugate of A. Hermitian matrices are therefore alternatively described as self-conjugate.

We proved above [32] that, for any vectors $\mathbf{x}$ and $\mathbf{y}$, an Hermitian matrix A satisfies

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A\mathbf{y}). \tag{183}$$

This relationship can serve like the two previous ones as a definition of Hermitian matrix.

The following further property should be noticed.

Let A be an Hermitian matrix and U any unitary matrix. Then we can easily show that $U^{-1}AU$ is likewise Hermitian. We have $\overline{A}^{(*)} = A$ by hypothesis. We want to show that $U^{-1}AU$ has the same property. From [26]:

$$\overline{(U^{-1}AU)}^{(*)} = \overline{U}^{(*)}\overline{A}^{(*)}\overline{U}^{(*)-1}$$

and this gives us, in view of the hypothesis for A and the unitary nature of U which implies $U^{(*)} = U^{-1}$:

$$\overline{(U^{-1}AU)}^{(*)} = U^{-1}AU$$

which is what we required to prove.

We can say that, for any unitary transformation of coordinates which is embodied for vector components in the expression

$$(x_1, \ldots, x_n) = U(x'_1, \ldots, x'_n),$$

an Hermitian A as operator of a linear transformation of space appears in the new coordinates as $U^{-1}AU$, so that the above proposition can also be stated as: unitary transformations of space do not change the Hermitian nature of a matrix as operator.

We now consider the problem of reducing an Hermitian matrix to the diagonal form with the aid of a unitary transformation:

$$U^{-1}AU = [\lambda_1, \ldots, \lambda_n]. \tag{184}$$

The problem is equivalent, as above in the case of real symmetric matrices, to the solution of an equation of the form

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{185}$$

where λ is one of the λ_k and the components of the vector $\mathbf{x}$ give the elements of the corresponding column of U.

The numbers λ_k and corresponding vectors $\mathbf{x}^{(k)}$ are known as the eigenvalues and eigenvectors of matrix A.

As we know, the eigenvalues must be the roots of the equation

$$\begin{vmatrix} a_{11}-\lambda, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22}-\lambda, \ldots, a_{2n} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{n1}, a_{n2}, \ldots, a_{nn}-\lambda \end{vmatrix} = 0. \tag{186}$$

Let $\lambda = \lambda_1$ be a root of this equation, and $\mathbf{x}^{(1)}$ be a solution of equation (185) with $\lambda = \lambda_1$.

Since (185) is linear and homogeneous, a solution can be multiplied by an arbitrary constant, and we can therefore take the length of $\mathbf{x}^{(1)}$ as unity. We take this vector as the first of the fundamental set in the new coordinates, then suitably complete the fundamental set with a further $(n - 1)$ orthogonal unit vectors. Let U_1 be the unitary transformation corresponding to passage to the new fundamental set. Our Hermitian A becomes the new Hermitian matrix $A_1 = U_1^{-1}AU_1$ in the new coordinates, whilst the corresponding equation

$$A_1\mathbf{x} = \lambda\mathbf{x}$$

will have the vector with components $(1, 0, \ldots, 0)$ as a solution with $\lambda = \lambda_1$. This fact shows us, as in [33], that all the elements of the first row and column of A_1 must vanish except the element λ_1 at their intersection.

It follows at once from the fact that A_1 is Hermitian that λ_1 must be real, and hence that every root of (186) must be real, as we saw above. The matrix A_1 can now be written as

$$\begin{Vmatrix} \lambda_1, 0, \ldots, 0 \\ 0, a_{22}^{(1)}, \ldots, a_{2n}^{(1)} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0, a_{n2}^{(1)}, \ldots, a_{nn}^{(1)} \end{Vmatrix},$$

i.e. it is a quasi-diagonal matrix of the form

$$[\lambda_1, C_1],$$

where C_1 denotes the Hermitian matrix of order $(n - 1)$ with elements $a_{ik}^{(1)}$. We can now repeat the above argument and reduce C_1 with the aid of a unitary transformation U_2 on all but the first of the

fundamental set to a form such that all the elements of the first row and column vanish except for the element at their intersection.

We can consider this latter unitary transformation as acting on the entire n-dimensional space and as given by the quasidiagonal unitary matrix

$$[1, U_2].$$

As a result of this transformation our Hermitian matrix becomes

$$[1, U_2]^{-1}[\lambda_1, C_1][1, U_2] = [\lambda_1, U_2^{-1} C_1 U_2],$$

and the new Hermitian matrix will have the expanded form

$$\begin{Vmatrix} \lambda_1, 0, & 0, \ldots, 0 \\ 0, \lambda_2, & 0, \ldots, 0 \\ 0, \ 0, & a_{33}^{(2)}, \ldots, a_{3n}^{(2)} \\ \cdots & \cdots \\ 0, \ 0, & a_{n3}^{(2)}, \ldots, a_{nn}^{(2)} \end{Vmatrix}.$$

By continuing in this way, we successively reduce our Hermitian matrix to the diagonal form, the total unitary transformation U appearing in (184) being the product of all the successive unitary transformations.

We return to equation (185). We proved in [33] that its solutions corresponding to the different values of λ must be mutually orthogonal.

We can show exactly as in [33] that the vectors formed by the columns of matrix U, together with the corresponding values of λ, yield all the solutions of (185). We only need to bear in mind here the following important fact regarding multiple roots of (186). If $\lambda = \lambda_1$ is say an m-tuple root of (186), (185) will have m linearly independent solutions $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$ for $\lambda = \lambda_1$. Every linear combination of these with arbitrary coefficients will obviously also be a solution of (185), i.e. the equation

$$A\mathbf{x} = \lambda_1 \mathbf{x}$$

has a set of solutions representing the subspace formed by the vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}$, or in other words, defined by the sum

$$x = C_1 \mathbf{x}^{(1)} + \ldots + C_m \mathbf{x}^{(m)}$$

with arbitrary coefficients $C_1, \ldots, C_m$. We can select in any manner a system of m mutually orthogonal unit vectors in this subspace, such that their components give the columns of the matrix U that

corresponds to the eigenvalue $\lambda = \lambda_1$. This means that we have here the same arbitrariness in the choice of U as we had in [33] for B. Moreover, we can obviously multiply the components of every $\mathbf{x}^{(s)}$, obtained by solving (185), by a numerical factor of unit modulus, i.e. by a factor of the form $e^{i\varphi}$ (the phase factor). The vector still retains its unit length after the multiplication, as well as its orthogonality to all the other vectors appearing in the complete system of solutions of (185). Finally, we can arbitrarily change the order of the columns in U. This is a trivial transformation that clearly amounts to re-numbering the fundamental set in the new coordinate system and merely involves a rearrangement of the λ_k in the diagonal matrix. We shall always assume in future that the λ_k are in increasing order.

We now turn to Hermitian forms. We shall say that the *Hermitian form*

$$A(\mathbf{x}) = (A\mathbf{x}, \mathbf{x}) = \sum_{i,\, k=1}^{n} a_{ik} \bar{x}_i x_k, \tag{187}$$

where $x_1, \ldots, x_n$ are the components of a vector $\mathbf{x}$, corresponds to the Hermitian matrix A. We have previously looked on matrix A as a linear transformation of space which yields a new vector $\mathbf{x}'$ on being applied to a given vector $\mathbf{x}$, and we have written the result of this transformation as $A\mathbf{x}$. In the expression $A(\mathbf{x})$, the final result is no longer a vector, but a number. We saw above that this number is real.

Now suppose we have carried out a unitary transformation of the space, the old vector components being given in terms of the new by $\mathbf{x} = U\mathbf{x}'$. The Hermitian form (187) becomes in the new coordinates:

$$(AU\mathbf{x}', U\mathbf{x}').$$

Property (125_1) of unitary transformations enables us to multiply both vectors in this scalar product on the left by the unitary matrix U^{-1}, so that we can now write for Hermitian form (187):

$$(U^{-1} AU\mathbf{x}', \mathbf{x}'). \tag{188}$$

In particular, if the unitary U transforms A to the diagonal form, i.e. (184) is valid, only the terms containing the products $\bar{x}_i' \cdot x_i'$ will remain in our Hermitian form in the new variables, and our form will have been reduced to a sum of squares:

$$(\mathbf{x}' \cdot U^{-1} AU\mathbf{x}) = \lambda_1 \bar{x}_1' x_1' + \lambda_2 \bar{x}_2' x_2' + \ldots + \lambda_n \bar{x}_n' x_n'.$$

Thus the task of transforming a matrix A to the diagonal form is equivalent here, as in [32], to the task of reducing the corresponding Hermitian form to a sum of squares.

Instead of Hermitian forms, *bilinear forms* are sometimes considered, these being defined by

$$(A\mathbf{y}, \mathbf{x}) = \sum_{i,\, k=1}^{n} a_{ik}\, \bar{x}_i\, y_k .$$

If we again apply a unitary transformation to the space so that the new components are given in terms of the old by the previous formula, we have in the new coordinates:

$$(A\mathbf{y}, \mathbf{x}) = (AU\mathbf{y}', U\mathbf{y}')$$

or, by the property of unitary transformations:

$$(U^{-1}AU\mathbf{y}', \mathbf{x}) .$$

Finally, if U reduces A to the diagonal form, the bilinear form reduces in the corresponding coordinates to the following simple form:

$$\sum_{k=1}^{n} \lambda_k\, \bar{x}'_k\, y'_k .$$

We notice that any diagonal matrix with real elements is Hermitian, so that $U^{-1}[\lambda_1, \ldots, \lambda_n]\, U$, where U is any unitary matrix, is also Hermitian. We saw above that, conversely, any Hermitian matrix can be written in this form.

Hermitian forms may be classified in the same way as real quadratic forms [35], according to the signs of the characteristic numbers λ_k. If all the λ_k are positive say, the Hermitian form is said to be positive definite. It is characterized in this case by being positive for any values of x_k and by vanishing only when $x_1 = \ldots = x_n = 0$. We can similarly define semi-definite and alternating Hermitian forms. The discussion follows exactly the same lines as for real quadratic forms and is based on the expression

$$(A\mathbf{x}, \mathbf{x}) = \lambda_1\, \bar{x}'_1 x'_1 + \ldots + \lambda_n\, \bar{x}'_n x'_n .$$

Equation (183) holds for Hermitian matrices. Given any matrix A and its conjugate $\tilde{A} = \overline{A}^{(*)}$, we have instead of (183):

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \tilde{A}\,\mathbf{y}) . \qquad (183_1)$$

If A has elements a_{ik}, $\tilde{A}$ has elements $\{\tilde{A}\}_{ik} = \bar{a}_{ki}$, and (183_1) may be verified by direct substitution as in the case of (183).

41. Commutative Hermitian matrices. Let A and B be two Hermitian matrices. We consider the conditions under which their product BA is likewise Hermitian. We write down the Hermitian conjugate of BA:

$$(\overline{BA})^{(*)} = \overline{A}^{(*)}\overline{B}^{(*)}$$

or, since A and B are Hermitian:

$$(\overline{BA})^{(*)} = AB .$$

The necessary and sufficient condition for BA to be Hermitian is for AB to coincide with BA, i.e. *for the matrices to commute.* Suppose that the Hermitian matrices A and B are reducible to the diagonal form by means of the same unitary transformation U:

$$A = U^{-1}[\lambda_1, \ldots, \lambda_n]\,U ; \quad B = U^{-1}[\mu_1, \ldots, \mu_n]\,U .$$

It can easily be seen that they commute in this case:

$$AB = BA = U^{-1}[\lambda_1\mu_1, \ldots, \lambda_n\mu_n]\,U .$$

We now prove the converse: *if two Hermitian matrices commute, they can be simultaneously reduced to the diagonal form with the aid of the same unitary transformation,* i.e. commutation of Hermitian matrices is not only a necessary, but also a sufficient condition for them to be reducible simultaneously with the aid of a unitary transformation to the diagonal form. Suppose, then, that $AB = BA$. We notice that similar matrices to these will also commute. For

$$(C^{-1}AC)(C^{-1}BC) = C^{-1}ABC = C^{-1}BAC,$$

and the same expression is found for the product

$$(C^{-1}BC)(C^{-1}AC) .$$

Suppose we choose for C a unitary transformation that reduces A to the diagonal form, and that we apply the same transformation to B. Since the new matrices commute, we can assume in the proof of our proposition that A in fact already has the diagonal form, i.e. its elements a_{ik} satisfy the condition

$$a_{ik} = 0 \quad \text{for} \quad i \neq k . \tag{189}$$

Let us denote the elements of B by b_{ik} and write down the condition that the matrices commute:

$$\sum_{s=1}^{n} a_{is} b_{sk} = \sum_{s=1}^{n} b_{is} a_{sk} \qquad (i, k = 1, 2, \ldots, n) .$$

for any i and k. This becomes, by (189):

$$(a_{ii} - a_{kk})\,b_{ik} = 0 \qquad (i,\, k = 1, 2, \ldots, n). \tag{190}$$

If all the a_{ii} are different, these last equations at once imply that $b_{ik} = 0$ for $i \neq k$, i.e. B is also a diagonal matrix, and the proposition is proved.

We now turn to the general case, when some of the a_{ii} are identical We may suppose for definiteness that they fall into two groups of equal numbers:

$$a_{11} = \ldots a_{mm}; \quad a_{m+1,\,m+1} = \ldots = a_{nn}.$$

It follows at once from (190) that in this case the b_{ik} can differ from zero only when either both subscripts i and k are greater than m, or when both are not greater than m. This means that B must here be quasi-diagonal:

$$B = [B_1, B_2],$$

where B_1 and B_2 are Hermitian matrices of orders m and $(n - m)$ respectively. We can write B out in full as

$$\left\| \begin{matrix} b_{11}, & \ldots, b_{mm}, & 0, & \ldots, 0 \\ \cdots & \cdots & \cdots & \cdots \\ b_{m1}, & \ldots, b_{mm}, & 0, & \ldots, 0 \\ 0, & \ldots, 0, & b_{m+1,m+1}, & \ldots, b_{m+1,\,n} \\ \cdots & \cdots & \cdots & \cdots \\ 0, & \ldots, 0, & b_{n,\,m+1}, & \ldots, b_{nn} \end{matrix} \right\|.$$

We can subject the subspace formed by the first m fundamental vectors to a unitary transformation without changing the diagonal form A, and the same is true for the subspace formed by the succeeding $(n - m)$ vectors. We choose these unitary transformations V_1 and V_2 so that B_1 and B_2 become diagonal. Altogether we have a unitary transformation of the n-dimensional space with the quasi-diagonal form

$$[V_1, V_2].$$

The matrix A remains diagonal in the new coordinates by what has been said above, while matrix B takes the form

$$[V_1, V_2]^{-1}[B_1, B_2][V_1, V_2] = [V_1^{-1} B_1 V_1, V_2^{-1} B_2 V_2],$$

i.e. is also diagonal, so that our proposition is proved.

If we now write the equations

$$Ax = \lambda x; \quad Bx = \mu x, \tag{191}$$

for our commuting matrices, it follows immediately from the above that we can form the same system of n linearly independent solutions for both the equations. These solutions in fact give the columns of the unitary matrix U that reduces both matrices to the diagonal form. In other words, we can form the same complete system of n linearly independent eigenvectors for two commuting Hermitian matrices. The eigenvalues, i.e. the values of the parameters λ and μ, are of course generally different. We remark that it does not follow from the above that every eigenvector of A is likewise an eigenvector of B. This would be the case, of course, if all the eigenvalues of A and B were distinct, so that a single vector, apart from a numerical factor, corresponded to each value λ_k and μ_k. But this is not generally true if some of the eigenvalues are equal. Let $\mathbf{x}^{(k)}$ be the total system of eigenvectors of matrices A and B, whilst λ_k and μ_k are the corresponding eigenvalues. Suppose, say, that $\lambda_1 = \lambda_2$, but $\mu_1 \neq \mu_2$. The vectors $C_1 \mathbf{x}^{(1)} + C_2 \mathbf{x}^{(2)}$ are now, for any choice of constants C_1 and C_2, eigenvectors of A but not of B.

The whole of the above discussion is easily carried over to the case of any number of matrices: given Hermitian matrices $A_1, \ldots, A_l$, the necessary and sufficient condition for them to be reducible simultaneously to the diagonal form with the aid of a unitary transformation is for them to commute in pairs, i.e. $A_i A_k = A_k A_i$ for any i and k from 1 to l.

42. The reduction of unitary matrices to the diagonal form. Unitary matrices have a similar property to Hermitian matrices as regards reduction to the diagonal form: if V is any unitary matrix, a unitary matrix U can always be found such that

$$U^{-1} V U$$

is diagonal. We can write the problem in the form

$$VU = U[\lambda_1, \ldots \lambda_k], \tag{192}$$

where U is a required unitary matrix and the λ_k are required numbers.

As in the earlier case of Hermitian matrices, the vectors $\mathbf{x}^{(k)}$ corresponding to the columns of U must be solutions of the equation

$$V\mathbf{x} = \lambda \mathbf{x}, \tag{193}$$

where λ is any of the λ_k. It follows at once from this, as above, that the λ must be the roots of the characteristic equation

$$\begin{vmatrix} v_{11}-\lambda, & v_{12}, & \ldots, & v_{1n} \\ v_{21}, & v_{22}-\lambda, & \ldots, & v_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ v_{n1}, & v_{r2}, & \ldots, & v_{nn}-\lambda \end{vmatrix} = 0, \tag{194}$$

where the elements of V are written v_{ik}.

We notice first of all that if V_1 and U_1 are unitary, $U^{-1} V_1 U_1$ is likewise unitary. For since U_1 is unitary, U_1^{-1} is unitary, and the product of unitary matrices is also unitary.

After substituting a root $\lambda = \lambda_1$ of equation (194) in (193) and finding the unit vector $\mathbf{x}^{(1)}$ satisfying (193), we take this as a new fundamental vector and associate with it a further $(n-1)$ unit vectors such that we have a system of n mutually orthogonal unit vectors. Passage from the old to the new fundamental set is equivalent to a unitary transformation U_1, and our unitary matrix V becomes the similar matrix

$$V_1 = U_1^{-1} V U_1 .$$

The equation

$$V_1 \mathbf{x} = \lambda \mathbf{x}$$

has the vector with components $(1, 0, \ldots, 0)$ as a solution for $\lambda = \lambda_1$, whence, as above, it follows immediately that the elements of the first column of V_1 are all zero except the first, which is equal to λ_1. But since, in a unitary matrix, the sum of the squares of the moduli of the elements of each column is unity, the number λ_1 can be said to have a modulus of unity. We now recall that, in the unitary matrix V_1, the sum of the squares of the moduli of the elements of each row must likewise be unity. But we have just shown that λ_1, the first element of the first row, has unit modulus, so that the remaining elements of the row must be zero. Thus our unitary transformation has reduced our unitary matrix to the form in which all elements except the first of the first row and column are zero:

$$\begin{Vmatrix} \lambda_1, & 0, & \ldots, & 0 \\ 0, & v_{22}^{(1)}, & \ldots, & v_{2n}^{(1)} \\ \ldots & \ldots & \ldots & \ldots \\ 0, & v_{n2}^{(1)}, & \ldots, & v_{nn}^{(1)} \end{Vmatrix}.$$

We had the same situation previously for Hermitian matrices. The elements $v_{ik}^{(1)}$ now form a unitary matrix of order $(n-1)$. We can

apply a further unitary transformation so as to obtain zeros in the first row and column of this matrix, except in the case of the first element, the modulus of which will be unity. As a final result of our two unitary transformations, the unitary matrix becomes

$$\begin{Vmatrix} \lambda_1, & 0, & 0, & \dots, & 0 \\ 0, & \lambda_2, & 0, & \dots, & 0 \\ 0, & 0, & v_{33}^{(2)}, & \dots, & v_{3n}^{(2)} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & v_{n3}^{(2)}, & \dots, & v_{nn}^{(2)} \end{Vmatrix}.$$

By continuing in this manner, our unitary matrix is reduced to the diagonal form with the aid of a certain unitary transformation. We remark that it follows at once from the above discussion that *all the characteristic roots of a unitary matrix have unit modulus.*

It can be shown as in [41] that if any number of unitary matrices commute in pairs, they can all be reduced to the diagonal form with the aid of the same unitary transformation.

We also notice the following point. Let a unitary matrix reduce a matrix A to the diagonal form, i.e. $U^{-1}AU$ is diagonal. We know that the modulus of the determinant of U is unity, so that we can find a real number ω such that the determinant of the unitary matrix $e^{i\omega}U$ is unity. But $e^{i\omega}U$ also reduces A to the diagonal form, since

$$(e^{i\omega}U)^{-1}A(e^{i\omega}U) = e^{i\omega}e^{-i\omega}U^{-1}AU = U^{-1}AU.$$

It follows that we can always take the determinant of a unitary matrix U reducing a given matrix to the diagonal form to be equal to unity.

Example. We take as an example the reduction to the diagonal form of a real third order orthogonal matrix:

$$V = \begin{Vmatrix} v_{11}, & v_{12}, & v_{13} \\ v_{21}, & v_{22}, & v_{23} \\ v_{31}, & v_{32}, & v_{33} \end{Vmatrix}. \tag{195}$$

We shall assume that the determinant of this matrix is equal to $(+1)$, so that the matrix corresponds to a movement about the origin of the three-dimensional space as a whole. The characteristic equation for matrix (195) has a constant term equal to unity by hypothesis, since the constant term evidently coincides with the determinant of the matrix. We have seen, on the other hand, that all the roots of the characteristic equation have unit modulus. The first term of the characteristic equation will be $(-\lambda)^3 = -\lambda^3$, and therefore the constant term viz., unity is identical to the product of the roots. Since the equation has real coef-

ficients, only two cases are possible: either one root is equal to 1, whilst the other two are imaginary conjugates of unit modulus, i.e. of the form $e^{\pm i\varphi}$, or else one root is 1, and the two others are (-1). The second case is a particular case of the first with $\varphi = \pi$.

The real vector $\mathbf{x}^{(1)}$, corresponding to the eigenvalue $\lambda = 1$, must be a solution of the equation

$$V\mathbf{x}^{(1)} = \mathbf{x}^{(1)}. \tag{196}$$

In other words, this vector must not change with the rotation of space defined by matrix V. The vector is real since it corresponds to the real value $\lambda = 1$, and it evidently defines the axis about which the space rotates (any rotation of space about the origin is equivalent to rotation about some axis through the origin). To find the components of $\mathbf{x}^{(1)}$ in terms of the elements of matrix V, we re-write (196) as

$$V^{-1}\mathbf{x}^{(1)} = \mathbf{x}^{(1)}$$

or, since V is real and unitary, we can write

$$V^{(*)}\mathbf{x}^{(1)} = \mathbf{x}^{(1)}.$$

We have on subtracting this from (196):

$$(V - V^{(*)})\,\mathbf{x}^{(1)} = 0.$$

We write this last equation out in full, the components of $\mathbf{x}^{(1)}$ being denoted by (u_{11}, u_{21}, u_{31}). This gives the system

$$\begin{aligned}
(v_{12} - v_{21})\,u_{21} + (v_{13} - v_{31})\,u_{31} &= 0\\
(v_{21} - v_{12})\,u_{11} \qquad\qquad + (v_{23} - v_{32})\,u_{31} &= 0\\
(v_{31} - v_{13})\,u_{11} + (v_{32} - v_{23})\,u_{21} \qquad\qquad &= 0,
\end{aligned}$$

whence the formula for the direction of the axis of rotation follows at once:

$$u_{11} : u_{21} : u_{31} = (v_{23} - v_{32}) : (v_{31} - v_{13}) : (v_{12} : - v_{21}).$$

The two other eigenvectors $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$ must clearly satisfy the equations

$$V\mathbf{x}^{(2)} = e^{i\varphi}\mathbf{x}^{(2)} \quad \text{and} \quad V\mathbf{x}^{(3)} = e^{-i\varphi}\mathbf{x}^{(3)}, \tag{197}$$

and these vectors now have complex components. We can find φ from the condition that the sum of the roots of the characteristic equation is evidently equal to the sum of the diagonal terms, i.e. to the trace of V:

$$1 + e^{-i\varphi} + e^{i\varphi} = 1 + 2\cos\varphi = v_{11} + v_{22} + v_{33},$$

where φ can be assumed to lie between 0 and π.

Since the values of λ in equations (197) are imaginary conjugates, it follows that we can assume that the components of $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(3)}$ are imaginary conjugates. We form the new unitary matrix

$$U_0 = \begin{Vmatrix} 1, & 0, & 0, \\ 0, & \dfrac{1}{\sqrt{2}}, & \dfrac{i}{\sqrt{2}} \\ 0, & \dfrac{1}{\sqrt{2}}, & -\dfrac{i}{\sqrt{2}} \end{Vmatrix}. \tag{198}$$

It may easily be verified directly that the elements of the columns of the matrix $W = UU_0$ are equal to the components of the vectors

$$\mathbf{x}^{(1)}; \qquad \frac{\mathbf{x}^{(2)} + \mathbf{x}^{(3)}}{\sqrt{2}}; \qquad i\,\frac{\mathbf{x}^{(2)} - \mathbf{x}^{(3)}}{\sqrt{2}},$$

i.e. they are real. Moreover W must also be unitary since it is the product of two unitary matrices, i.e. W is orthogonal. We now use the real unitary matrix W to apply a similarity transformation to V. This gives

$$W^{-1}VW = U_0^{-1}U^{-1}VUU_0 = U_0^{-1}[1, e^{i\varphi}, e^{-i\varphi}]\,U_0.$$

On carrying out the actual matrix multiplication, we get

$$W^{-1}VW = \begin{Vmatrix} 1, & 0, & 0 \\ 0, & \cos\varphi, & -\sin\varphi \\ 0, & \sin\varphi, & \cos\varphi \end{Vmatrix}. \tag{199}$$

We can always suppose that the determinant of the orthogonal matrix W is $(+1)$, since we could multiply the matrix by (-1) if this were not the case, as a result of which (199) would remain unchanged. Hence W will also correspond to a rotation of three-dimensional space. Matrix (199), obtained as a result of the coordinate transformation $\mathbf{x}' = W\mathbf{x}$, is similar to V and yields the same transformation in the new coordinates as the original matrix V gave in the old. It follows directly from the form of matrix (199) that this corresponds to a rotation about a new axis $\mathbf{x}'^{(2)}$ by an angle φ, and the essence of our transformation amounts to our having used as axis $\mathbf{x}'^{(1)}$ the above-mentioned axis of rotation represented by the vector $\mathbf{x}^{(1)}$.

A further important fact follows at once from the above: all the real matrices corresponding to a rotation of space by a given angle φ can be reduced to the same form (199) with the aid of a similarity transformation (different for different matrices), so that such matrices are similar to each other.

The matrices corresponding to different angles of rotation cannot be similar, since the different values of φ lead to different sets of characteristic roots 1, $e^{i\varphi}$, $e^{-i\varphi}$. All these properties have an extremely simple geometrical interpretation.

43. Projection matrices. We shall now consider a particular case of Hermitian matrices. Let R_m be the m-dimensional subspace formed by the linearly independent vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(m)}$. The subspace R_m consists of the set of vectors of the form

$$C_1\mathbf{y}^{(1)} + \ldots + C_m\mathbf{y}^{(m)},$$

where the C_k are arbitrary numerical coefficients. We can orthogonalize the $\mathbf{y}^{(k)}$ and form m mutually orthogonal unit vectors

$$\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)},$$

which yield the same subspace R_m. Then we can make these into a complete system of n mutually orthogonal unit vectors by constructing a further $(n - m)$ unit vectors

$$\mathbf{x}^{(m+1)}, \ldots, \mathbf{x}^{(n)}.$$

These last vectors form an $(n-m)$-dimensional subspace R'_{n-m}, the two subspaces R_m and R_{n-m} being mutually orthogonal in the sense that any vector of the former is orthogonal to any vector of the latter [14]. On writing any vector $\mathbf{x}$ in terms of the fundamental set $\mathbf{x}^{(k)}$:

$$\mathbf{x} = x_1\mathbf{x}^{(1)} + \ldots + x_n\mathbf{x}^{(n)}, \tag{200}$$

we can represent it as the sum of two vectors:

$$\mathbf{x} = [x_1\mathbf{x}^{(1)} + \ldots + x_m\mathbf{x}^{(m)}] + [x_{m+1}\mathbf{x}^{(m+1)} + \ldots + x_n\mathbf{x}^{(n)} = \mathbf{u} + \mathbf{v}, \tag{201}$$

one of which belongs to R_m and the other to R'_{n-m}. It is easily seen that this resolution of any vector $\mathbf{x}$ into two components is unique. For suppose that, in addition to (201), we have a second resolution $\mathbf{x} = \mathbf{u}' + \mathbf{v}'$ with the above property. Then

$$\mathbf{u} + \mathbf{v} = \mathbf{u}' + \mathbf{v}' \quad \text{or} \quad \mathbf{u} - \mathbf{u}' = \mathbf{v}' - \mathbf{v}.$$

The vector on the left belongs to R_m and that on the right to R'_{n-m}, so that $\mathbf{u} - \mathbf{u}'$ and $\mathbf{v} - \mathbf{v}'$ are orthogonal.

But any vector orthogonal to itself is clearly zero [14], and consequently $\mathbf{u} - \mathbf{u}' = 0$, i.e. $\mathbf{u}$ is the same as $\mathbf{u}'$, whilst $\mathbf{v}$ is the same as $\mathbf{v}$, i.e. $\mathbf{u}$ and $\mathbf{v}$ are uniquely defined for $\mathbf{x}$. The vector $\mathbf{u}$ is called *the projection of $\mathbf{x}$ on the subspace R_m*. The matrix for passing from $\mathbf{x}$ to $\mathbf{u}$ is called the *projection matrix* on to the subspace R_m and is written P_{R_m}. The form of this matrix naturally depends on the choice of coordinate axes.

If we take the $\mathbf{x}^{(k)}$ as fundamental set, $\mathbf{x}$ is given by (201), whilst $\mathbf{u}$ is given by

$$\mathbf{u} = x_1\mathbf{x}^{(1)} + \ldots + x_m\mathbf{x}^{(m)},$$

and the operation of projection here amounts simply to leaving the first m components as before and putting the remainder equal to zero. The corresponding projection matrix is clearly diagonal:

$$P_{R_m} = [1, 1, \ldots, 1, 0, 0, \ldots, 0],$$

where we have unity in the first m places and zero elsewhere. If the fundamental set were numbered differently, we should still get a diagonal matrix of ones and zeros, though in a different order. In the general case of any choice of Cartesian axes, the projection matrix has the form:

$$P_{R_m} = U^{-1}[1, \ldots, 1, 0, \ldots, 0]\,U, \tag{202}$$

where U is unitary, and the eigenvalues of P_{R_m} are either zero or unity. Conversely, every Hermitian matrix of this form is a projection matrix on to a subspace whose number of dimensions is given by the number of eigenvalues of P_{R_m} equal to unity.

A projection matrix can be alternatively defined as follows: *a projection matrix is an Hermitian matrix satisfying the equation*

$$P^2 = P. \tag{203}$$

For we can easily verify that a matrix of the form (202) satisfies relationship (203) on noticing that $1^2 = 1$ and $0^2 = 0$. Conversely, if an Hermitian matrix satisfies (203), and we write it

$$P = U^{-1}[\lambda_1, \dots \lambda_n]\,U,$$

we have by (203):

$$U^{-1}[\lambda_1^2, \dots, \lambda_n^2]\,U = U^{-1}[\lambda_1 \dots, \lambda_n]\,U,$$

i.e. $\lambda_k^2 = \lambda_k$ $(k = 1, 2, \dots, n)$, whence it follows at once that λ_k is unity or zero. If all the characteristic roots of the matrix are unity, we have the unit matrix which corresponds to the identity transformation; in other words, a vector is projected on to the total space (and remains unchanged). Apart from this trivial case, we have at least one zero characteristic root in the projection matrix, so that the determinant of the matrix, equal to the product of the characteristic roots, is also zero, i.e. there is no question of an inverse matrix P^{-1}. We notice that it also follows directly from the definition that the projection matrix P_{R_m} does not change a vector belonging to the subspace R_m, and diminishes the length of a vector not belonging to R_m.

We follow these preliminary observations by considering some operations with projection matrices. Let us have two projection matrices P_R and P_S such that their product is zero, i.e. all the elements of the product matrix are zero:

$$P_S P_R = 0. \tag{204}$$

We take a vector $\mathbf{x}$ of the subspace R, such that $P_R\,\mathbf{x} = \mathbf{x}$. Equation (204) gives us

$$P_S\,\mathbf{x} = 0.$$

But it follows directly from this that $\mathbf{x}$ is orthogonal to any vector of the subspace S. For otherwise we could find a unit vector $\mathbf{y}$ of S not orthogonal to $\mathbf{x}$ and on taking this as the first of the fundamental set, we should have a non-zero magnitude for the first component of $\mathbf{x}$ which would remain unchanged on projection of $\mathbf{x}$ on to S. Hence we see that, if condition (204) is satisfied, every vector of R is orthogonal to every vector of S, and conversely. We now have, in addition to (204):

$$P_R\,P_S = 0. \tag{205}$$

For, given any vector $\mathbf{y}$, the vector $P_S\,\mathbf{y}$ belongs to S and is therefore orthogonal to every vector of R, i.e. we have for any $\mathbf{y}$:

$$P_R\,P_S \mathbf{y} = 0,$$

which is equivalent to (205). Conversely, if two subspaces R and S are orthogonal in the above sense, (204) and (205) are valid.

We now consider the sum of the projection matrices:

$$P = P_R + P_S \tag{206}$$

and assume that (204) and (205) are satisfied. We show that (206), which is clearly Hermitian, is also a projection matrix, i.e. we want to show that it is equal to its square:

$$P^2 = (P_R + P_S)\,(P_R + P_S) = P_R^2 + P_R P_S + P_S P_R + P_S^2,$$

from which we have, in fact, in view of our conditions and the fact that P_R and P_S are projection matrices:

$$P^2 = P_R + P_S = P.$$

It may easily be seen that P corresponds in the present case to a projection on to the subspace $(R + S)$, where the addition of R and S is taken to mean the subspace consisting of all the vectors which are the sums of vectors of R and of S, i.e. if R is formed from $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p)}$, and S from $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(q)}$, $(R + S)$ consists of the system

$$C_1\mathbf{x}^{(1)} + \ldots + C_p\mathbf{x}^{(p)} + D_1\mathbf{y}^{(1)} + \ldots + D_q\mathbf{y}^{(q)},$$

where the C_k and D_k are arbitrary constants. The above property may be generalized for any number of terms:

$$P = P_{S_1} + \ldots + P_{S_m}. \tag{207}$$

If the subspaces S_k are orthogonal in pairs, i.e. any vector of S_i is orthogonal to any vector of S_j for differing i and j, sum (207) represents the projection matrix on to the subspace $(S_1 + \ldots + S_m)$, formed from all the vectors used for forming the S_k. In particular, the sum can be equal to the unit matrix:

$$I = P_{S_1} + \ldots + P_{S_m},$$

and we usually speak in this case of the resolution of the identity into projection matrices, or simply, of the resolution of the identity.

We next consider the product of two projection matrices:

$$P = P_S P_R. \tag{208}$$

For the product to be likewise a projection matrix, we first of all require it to be Hermitian which implies in turn [41] that the matrices commute:

$$P_R P_S = P_S P_R. \tag{209}$$

This condition may be shown to be sufficient, i.e. $P^2 = P$ in this case. We have

$$P^2 = P_S P_R P_S P_R$$

or, on commuting the matrices in accordance with (209):

$$P^2 = P_S^2 P_R^2 = P_S P_R$$

which is what we required to prove. It may easily be verified that, given commutation condition (209), matrix (208) corresponds to the subspace formed by the vectors common to the two sets that form R and S.

We also notice a simple result, the proof of which we need not dwell on: if S forms part of subspace R, the difference

$$P = P_R - P_S \tag{210}$$

is also a projection matrix. If we take $\mathbf{x}^{(k)}$ as the fundamental set forming S, we have to add one or more linearly independent vectors in order to get the fundamental set forming R. These added vectors themselves form a subspace T, and the projection matrix on to T is given by matrix (210).

By using projection matrices, we can state the problem of reducing a Hermitian matrix to the diagonal form in a precise manner even with the presence of multiple eigenvalues.

Suppose, for instance, that we have the Hermitian matrix

$$A = U\,[\lambda_1, \ldots, \lambda_n]\,U^{-1},$$

where U is a unitary matrix. Suppose for definiteness that the λ_k fall into two groups, the m of the first group being all equal to μ, and the remaining $(n - m)$ of the second group being equal to ν:

$$A = U\,[\mu, \ldots, \mu, \nu, \ldots, \nu]\,U^{-1}.$$

We can evidently re-write our matrix as

$$A = \mu U\,[1, \ldots, 1, 0, \ldots, 0]\,U^{-1} + \nu\,U\,[0, \ldots, 0, 1, \ldots, 1]\,U^{-1}.$$

We now introduce the projection matrices

$$P_R = U\,[1, \ldots, 1, 0, \ldots, 0]\,U^{-1};\ P_S = U\,[0, \ldots, 0, 1, \ldots, 1]\,U^{-1}.$$

The corresponding subspaces R and S are obviously orthogonal, and addition of the projection matrices yields a unique matrix. We thus have

$$A = \mu P_R + \nu P_S,$$

where

$$\lambda_1 = \ldots = \lambda_m = \mu \text{ and } \lambda_{m+1} = \ldots = \lambda_n = \nu.$$

The problem of reducing an Hermitian matrix to the diagonal form amounts in general to a resolution of the indentity

$$I = P_{S_1} + \ldots + P_{S_m}, \tag{211}$$

such that A is expressible in the form

$$A = \mu_1 P_{S_1} + \cdots + \mu_m P_{S_m}, \tag{212}$$

where the μ_k are the different eigenvalues of our matrix A. Thus to every Hermitian matrix there corresponds a definite resolution of the identity (211) such that the matrix is expressible in the form (212).

All the above results can easily be translated into the language of Hermitian forms instead of matrices. For every projection matrix P_R with elements p_{ik} we have a corresponding Hermitian form:

$$P_R(x) = (P_R x, x) = \sum_{i,\,k=1}^{n} p_{ik}\bar{x}_i x_k \tag{213}$$

which is sometimes called an *Einzelform*. If the corresponding subspace R has m dimensions, and we take m mutually orthogonal unit vectors of R as the first m of our fundamental set, form (213) becomes in this coordinate system:

$$(P_R \mathbf{x}', \mathbf{x}') = \bar{x}_1' x_1' + \bar{x}_2' x_2' + \ldots + \bar{x}_m' x_m'.$$

We observe further that, if the matrices P_{S_k} are the resolution of the identity given by (211), we clearly have, on choosing as fundamental set mutually

orthogonal unit vectors from each of the subspaces S_k:

$$\sum_{k=1}^{m} P_{S_k}(\mathbf{x}') = \sum_{i=1}^{n} x_i'\bar{x}_i',$$

and consequently the sum

$$\sum_{k=1}^{m} P_{S_k}(\mathbf{x})$$

gives the square of the length of the vector for any choice of coordinate axes. We can therefore say that *the problem of reducing an Hermitian form A to a sum of squares is equivalent to solving the two equations:*

$$A(\boldsymbol{x}) = \sum_{k=1}^{m} \mu_k P_{S_k}(\mathbf{x}), \tag{214}$$

$$|\boldsymbol{x}|^2 = \sum_{k=1}^{m} P_{S_k}(\mathbf{x}). \tag{215}$$

The introduction of the projection matrices thus allows of a statement of the problem of reducing an Hermitian matrix to the diagonal form without any special choice of coordinate axes. This in turn makes it possible to extend the above results, with suitable changes, to the case of space with an infinity of dimensions, which is the basic mathematical problem from the point of view of present-day quantum mechanics. We shall not discuss this till later. This extension to the case of an infinite set of dimensions carries us outside the realm of algebra and is intimately connected with the introduction of the apparatus of analysis.

44. Functions of matrices. Matrices can take the role of the arguments of functions. We confine ourselves here to considering the most elementary functions, viz., *matrix polynomials and rational fractions.* A more detailed treatment of the theory of functions of matrices will be found later, after the theory of functions of a complex variable. A polynomial $f(A)$ of degree m in the variable matrix A has the form:

$$f(A) = c_0 + c_1 A + \ldots + c_m A^m, \tag{216}$$

where the c_k are numerical coefficients. The value of the function is given by the matrix whose elements are evidently

$$\{f(A)\}_{ik} = c_0\delta_{ik} + c_1\{A\}_{ik} + \ldots + c_m\{A^m\}_{ik},$$

where

$$\delta_{ik} = 0 \text{ for } i \neq k \text{ and } \delta_{ii} = 1.$$

We can also consider a polynomial in several matrices but have to bear in mind the non-commutativeness of matrices on multiplication.

A second degree polynomial in two variable matrices A and B has the general form

$$f(A, B) = c_0 + c_1 A + c_2 B + c_3 A^2 + c_4 B^2 + c_5 AB + c_6 BA .$$

We replace the A in (216) by the similar matrix $U^{-1} AU$. We have, on recalling that $(U^{-1} AU)^k = U^{-1} A^k U$,

$$f(U^{-1} AU) = c_0 + c_1 U^{-1} AU + \ldots + c_m U^{-1} A^m U = \\ = U^{-1}(c_0 + c_1 A + \ldots + c_m A^m) U ,$$

i.e.

$$f(U^{-1} AU) = U^{-1} f(A) U . \tag{217}$$

An analogous expression holds for a polynomial in several matrices:

$$f(U^{-1} AU, U^{-1} BU) = U^{-1} f(A, B) U . \tag{218}$$

We next dwell in rather more detail on the case of Hermitian matrices. If A is Hermitian, it follows directly from the definition that any positive integral power A^k, and the product cA, where c is a real constant, are also Hermitian. Moreover, the sum of Hermitian matrices is Hermitian. Hence it follows directly that if A is Hermitian in (216) and the coefficients c_k are real numbers, the value of the function $f(A)$ is also Hermitian. The Hermitian matrix $f(A)$ clearly commutes with A, and they can be simultaneously reduced to the diagonal form with the aid of some unitary transformation. We notice firstly that if we substitute a diagonal matrix $[\lambda_1, \ldots, \lambda_n]$ for A in function (216), we clearly get another diagonal matrix:

$$\sum_{k=0}^{m} c_k [\lambda_1^k, \ldots, \lambda_n^k] = [f(\lambda_1), \ldots, f(\lambda_n)] , \tag{219}$$

where $f(\lambda_k)$ is the numerical value of the polynomial on substituting the numbers λ_k for A.

Now let V be the unitary transformation reducing A to the diagonal form:

$$A = V [\lambda_1, \ldots, \lambda_n] V^{-1} .$$

We have by (217) and (219):

$$f(A) = V [f(\lambda_1), \ldots, f(\lambda_n)] V^{-1} ,$$

i.e. V also reduces $f(A)$ to the diagonal form, the eigenvalues of this last being $f(\lambda_k)$.

We now turn to rational fractions. Let $f_1(A)$ and $f_2(A)$ be two polynomials in the matrix A. We consider their quotient:

$$\frac{f_1(A)}{f_2(A)}. \tag{220}$$

We saw earlier [26] that the quotient of two matrices does not in general have a definite meaning. It may be shown in the present case, however, that (220) has a definite value provided only that the determinant of matrix $f_2(A)$ differs from zero. We can write (220) in two ways:

$$f_1(A)\,f_2(A)^{-1} \text{ or } f_2(A)^{-1} f_1(A).$$

We shall show that these two expressions are equal:

$$f_1(A)\,f_2(A)^{-1} = f_2(A)^{-1} f_1(A),$$

or what amounts to the same thing:

$$f_2(A)\,f_1(A) = f_1(A)\,f_2(A). \tag{221}$$

Since our polynomials contain only the single matrix A, they commute, i.e. (221) in fact holds, and quotient (220) has a single value. It is easily shown further that, in the case of a single matrix, rational fractions can be multiplied like ordinary fractions. For

$$\frac{f_1(A)}{f_2(A)} \cdot \frac{f_3(A)}{f_4(A)} = f_1(A)\,f_2(A)^{-1} f_3(A)\,f_4(A)^{-1},$$

or, since the terms commute:

$$\frac{f_1(A)}{f_2(A)} \cdot \frac{f_3(A)}{f_4(A)} = f_1(A)\,f_3(A) \cdot [f_2(A)\,f_4(A)]^{-1} = \frac{f_1(A)\,f_3(A)}{f_2(A)\,f_4(A)}.$$

We take as an example the rational fraction

$$U = \frac{1 + iA}{1 - iA}, \tag{222}$$

where A is an Hermitian matrix, i.e. $\bar{A}^{(*)} = A$. It is easy to see that U is unitary, i.e.

$$\bar{U}^{(*)} = U^{-1}. \tag{223}$$

For we have

$$\bar{U} = \frac{1 - i\bar{A}}{1 + i\bar{A}} = (1 - i\bar{A})(1 + i\bar{A})^{-1},$$

whence we get, on passing to the transposed matrix [26]:

$$\bar{U}^{(*)} = (1 + i\bar{A})^{(*)-1}(1 - i\bar{A})^{(*)} = (1 + i\bar{A}^{(*)})^{-1}(1 - i\bar{A}^{(*)}),$$

or, since $\bar{A}^{(*)} = A$:

$$\bar{U}^{(*)} = (1 + iA)^{-1}(1 - iA) = \frac{1 - iA}{1 + iA} = U^{-1},$$

i.e. (223) is satisfied, and U is in fact unitary.

We can write (222) in the form

$$U(1 - iA) = (1 + iA),$$

whilst the fact that U commutes with A by (223) means that

$$A = i\,\frac{U - 1}{U + 1}. \tag{224}$$

It can be shown, precisely as above, that if U is unitary and the determinant of the matrix $U + 1$ differs from zero, the matrix A defined by (224) is Hermitian. Hence any unitary matrix for which $D(U + 1) \neq 0$ can be written in terms of an Hermitian matrix A in accordance with (222).

45. Infinite-dimensional space. We now set about introducing the concept of space with an infinite set of dimensions. We need the preliminary idea of the limit of a complex variable. Let the complex variable $z = x + yi$ take the sequence of values:

$$z_1 = x_1 + y_1 i; \quad z_2 = x_2 + y_2 i; \ldots; \quad z_n = x_n + y_n i; \ldots \tag{225}$$

We say that the complex number $a = a + bi$ is the limit of sequence (225) if the modulus of the difference $(a - z_n)$ tends to zero on indefinite increase of n, i.e. $|a - z_n| \to 0$ as $n \to \infty$, and we write $a = \lim z_n$ or $z_n \to a$. But

$$|a - z_n| = |(a - x_n) + (b - y_n)i| = \sqrt{(a - x_n)^2 + (b - y_n)^2}.$$

Since both terms under the radical are non-negative, the condition $|a - z_n| \to 0$ is equivalent to the two conditions: $x_n \to a$ and $y_n \to b$. Thus

$$x_n + y_n i \to a + bi \tag{226}$$

is equivalent to $x_n \to a$ and $y_n \to b$. We consider the complex series:

$$\sum_{k=1}^{\infty} (a_k + b_k i). \tag{227}$$

It is said to be convergent if the sum of its first n terms:

$$S_n = \sum_{k=1}^{n} (a_k + b_k i) = (a_1 + a_2 + \ldots + a_n) + (b_1 + b_2 + \ldots + b_n)i$$

tends to a limit: $S_n \to a + bi$ on indefinite incrase of n, the limit $(a + bi)$ being called the sum of the series. It follows from this definition that the convergence of series (227) is equivalent to the convergence of the series

$$a = \sum_{k=1}^{\infty} a_k \text{ and } b = \sum_{k=1}^{\infty} b_k, \tag{228}$$

formed from the real and imaginary parts of the terms of (227).

Suppose that the series

$$\sum_{k=1}^{\infty} |a_k + ib_k| = \sum_{k=1}^{\infty} \sqrt{a_k^2 + b_k^2}, \tag{229}$$

formed from the moduli of the terms of (227), is convergent. In view of the obvious inequalities

$$|a_k| \leqslant \sqrt{a_k^2 + b_k^2} \quad \text{and} \quad |b_k| \leqslant \sqrt{a_k^2 + b_k^2}, \tag{230}$$

series (228) will now also be convergent and in fact converge absolutely, and series (227) is therefore also convergent, i.e. if series (229) is convergent, (227) is certainly convergent. Series (227) is said to be *absolutely convergent* in this case. By applying the usual Cauchy test, we can state the necessary and sufficient condition for absolute convergence as follows: given any small positive ε, there exists an N such that

$$\sum_{k=n}^{n+p} |a_k + ib_k| < \varepsilon, \tag{231}$$

where p is any positive integer and $n > N$.

We now apply the above to some particular cases that are essential to what follows. We take the series

$$\sum_{k=1}^{\infty} \alpha_k \beta_k, \tag{232}$$

where α_k and β_k are complex numbers, and where it is known that the series

$$\sum_{k=1}^{\infty} |\alpha_k|^2 \text{ and } \sum_{k=1}^{\infty} |\beta_k|^2 \tag{233}$$

are convergent. We use the inequality proved in [29]:

$$\left\{\sum_{k=n}^{n+p} |\alpha_k \beta_k|\right\}^2 \leqslant \sum_{k=n}^{n+p} |\alpha_k|^2 \sum_{k=n}^{n+p} |\beta_k|^2 .$$

We obtain from this, on taking into account the convergence of series (233), the fact that the sum

$$\sum_{k=n}^{n+p} |a_k \beta_k|$$

is as small as we please for large n and any p, i.e. the convergence of series (233) guarantees the absolute convergence of series (232).

We now consider

$$\sum_{k=1}^{\infty} |a_k + \beta_k|^2 = \sum_{k=1}^{\infty} (a_k + \beta_k)(\overline{a}_k + \overline{\beta}_k), \tag{234}$$

series (233) being assumed convergent as before. Series (234) can be written as the sum of four series:

$$\sum_{k=1}^{\infty} |a_k|^2; \quad \sum_{k=1}^{\infty} |\beta_k|^2; \quad \sum_{k=1}^{\infty} a_k \overline{\beta}_k; \quad \sum_{k=1}^{\infty} \overline{a}_k \beta_k.$$

The first two are convergent by hypothesis, whilst the last two are convergent in view of the proposition proved above, i.e. the convergence of series (233) implies the convergence of (234).

We now turn to *space with an infinite number of dimensions*. A *vector in this space* is defined by an infinite sequence of complex numbers:

$$\mathbf{x}(x_1, x_2, \ldots),$$

these numbers being always assumed subject to the condition that the series

$$\sum_{k=1}^{\infty} |x_k|^2 \tag{235}$$

is convergent. The aggregate of such vectors is generally called *Hilbert space*, the first investigation of this space being due to Hilbert. In future we shall write H for the space for brevity.

As above, we bring in the basic operations of multiplication of a vector by a number and addition of vectors for vectors of the space H. If we write the components of $\mathbf{x}$ as x_k, we take the components of $c\mathbf{x}$, where c is a complex number, as equal to cx_k. If x_k and y_k are the components of $\mathbf{x}$ and $\mathbf{y}$, the components of the vector $(\mathbf{x} + \mathbf{y})$ are taken to be equal to $(x_k + y_k)$. The difference $\mathbf{x} - \mathbf{y}$ is the sum

of $\mathbf{x}$ and $(-1)\mathbf{y}$ (cf. [12]). Since series (235) is convergent, the series $\sum_{k=1}^{\infty} | cx_k |^2$ is also convergent. Similarly, if

$$\sum_{k=1}^{\infty} | x_k |^2 \text{ and } \sum_{k=1}^{\infty} | y_k |^2$$

are convergent, it follows from what has been said above that

$$\sum_{k=1}^{\infty} | x_k + y_k |^2$$

is also convergent, i.e. the numerical sequences $(cx_1, cx_2, \ldots)$ and $(x_1 + y_1, x_2 + y_2, \ldots)$ define the vectors $c\mathbf{x}$ and $\mathbf{x} + \mathbf{y}$ in H, if $\mathbf{x}$ and $\mathbf{y}$ belong to H. The null vector is the vector, all the components of which are zero. It is simply denoted by the number 0 in vector equations.

Operations on the vectors are subject to the usual rules (cf. [12]):

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}; \quad (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z});$$
$$(a + b)\,\mathbf{x} = a\mathbf{x} + b\mathbf{x}; \quad a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}; \quad a(b\mathbf{x}) = (ab)\,\mathbf{x}.$$

Similarly, from what has been said, we can define the scalar product of two vectors $\mathbf{x}$ and $\mathbf{y}$ of the space:

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\infty} x_k \overline{y}_k .$$

The sum

$$(\mathbf{x}, \mathbf{x}) = \sum_{k=1}^{\infty} | x_k |^2 \tag{236}$$

defines the square of the length, or in other words, the norm of the vector $\mathbf{x}$. We introduce the following notation for this:

$$\sum_{k=1}^{\infty} | x_k |^2 = \| \mathbf{x} \|^2 . \tag{237}$$

The norm of any vector is positive, except in the case of the null vector, the norm of which is zero. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be orthogonal if their scalar product is zero, i.e. $(\mathbf{u}, \mathbf{v}) = 0$ or $(\mathbf{v}, \mathbf{u}) = 0$, one equation being a consequence of the other. Scalar products are subject to the same fundamental laws as in the case of a finite number of dimensions [13 and 30]. In particular we have the inequality

$$| (\mathbf{x}, \mathbf{y}) | \leqslant \| \mathbf{x} \| \cdot \| \mathbf{y} \| , \tag{238}$$

and the triangle rule follows exactly as in [30]:

$$\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|. \tag{239}$$

If the vectors $\mathbf{x}^{(k)}$ $(k = 1, 2, \ldots, m)$ are orthogonal in pairs, i.e. $(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = 0$ for $i \neq j$, we obviously have

$$(\mathbf{x}^{(1)} + \ldots + \mathbf{x}^{(m)}, \mathbf{x}^{(1)} + \ldots + \mathbf{x}^{(m)}) = (\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) + \ldots + (\mathbf{x}^{(m)}, \mathbf{x}^{(m)}),$$

or, what amounts to the same thing:

$$\|\mathbf{x}^{(1)} + \ldots + \mathbf{x}^{(m)}\|^2 = \|\mathbf{x}^{(1)}\|^2 + \ldots + \|\mathbf{x}^{(m)}\|^2, \tag{240}$$

i.e. *the square of the norm of a sum of vectors that are orthogonal in pairs is equal to the sum of the squares of the norms of the terms.* This proposition may be termed *Pythagoras' theorem*. It follows at once from the definition of norm that, if c is a complex number, we have for the norm of $c\mathbf{x}$:

$$\|c\mathbf{x}\| = |c\mathbf{x}| \cdot \|\mathbf{x}\|.$$

If the vectors $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \ldots, \mathbf{z}^{(m)}$ are orthogonal in pairs and the norm of each is equal to unity, i.e.

$$(\mathbf{z}^{(p)}, \mathbf{z}^{(q)}) = 0 \quad \text{for} \quad p \neq q,$$
$$(\mathbf{z}^{(p)}, \mathbf{z}^{(q)}) = 1,$$

(240) gives

$$\|c_1\mathbf{z}^{(1)} + \ldots + c_m\mathbf{z}^{(m)}\|^2 = |c_1|^2 + \ldots + |c_m|^2,$$

where the c_s are arbitrary complex numbers.

The fundamental vectors in our space H have the components

$$\mathbf{a}^{(1)}(1,\ 0,\ 0,\ \ldots); \quad \mathbf{a}^{(2)}(0,\ 1,\ 0\ \ldots); \ldots$$

The $\mathbf{a}^{(k)}$ are mutually orthogonal unit vectors. We can write the components x_k of the vector $\mathbf{x}$ as scalar products:

$$x_k = (\mathbf{x}, \mathbf{a}^{(k)}).$$

We again consider an arbitrary system of m mutually orthogonal vectors of unit length

$$\mathbf{z}^{(k)} \ (k = 1,\ 2, \ldots,\ m).$$

The scalar product $(\mathbf{x}, \mathbf{z}^{(k)})$ *defines the magnitude of the projection of* $\mathbf{x}$ *on the axis* $\mathbf{z}^{(k)}$. The $\mathbf{z}^{(k)}$ do not form a complete system of axes for the space H, and the sum

$$\sum_{k=1}^{m} (\mathbf{x}, \mathbf{z}_k)\, \mathbf{z}^{(k)}$$

in general differs from $\mathbf{x}$. We can write our vector $\mathbf{x}$ as:

$$\mathbf{x} = \sum_{k=1}^{m} (\mathbf{x}, \mathbf{z}^{(k)})\, \mathbf{z}^{(k)} + \mathbf{u}. \tag{241}$$

On forming the scalar product on the right of both sides with $\mathbf{z}^{(l)}$ and recalling that the $\mathbf{z}^{(k)}$ are mutually orthogonal unit vectors, we get

$$(\mathbf{x},\ \mathbf{z}^{(l)}) = (\mathbf{x},\ \mathbf{z}^{(l)}) + (\mathbf{u},\ \mathbf{z}^{(l)}),$$

i.e. $(\mathbf{u}, \mathbf{z}^{(l)}) = 0$, or in other words, $\mathbf{u}$ is orthogonal to all the $\mathbf{z}^{(k)}$. We can thus apply Pythagoras' theorem to the sum (241):

$$\|\mathbf{x}\|^2 = \sum_{k=1}^{m} |(\mathbf{x},\ \mathbf{z}^{(k)})|^2 + |\mathbf{u}|^2,$$

whence the so-called *Bessel inequality* follows at once:

$$\|\mathbf{x}\|^2 \geqslant \sum_{k=1}^{m} |(\mathbf{x}, \mathbf{z}^{(k)})|^2. \tag{242}$$

This can be stated as follows: *the sum of the squares of the moduli of the projections of a vector on to any given mutually orthogonal unit vectors is not greater than the square of the length (norm) of the projected vector itself.* We have the sign of equality in (242) when and only when the vector $\mathbf{u}$ in (241) is zero, i.e. its components are all zero.

46. The convergence of vectors. We now explain the idea of the *limit of a variable vector*. Suppose we have a sequence of vectors $\mathbf{v}^{(k)}$, where k takes the values 1, 2, 3, ...

We denote the components of $\mathbf{v}^{(k)}$ by $v_1^{(k)}, v_2^{(k)}, \ldots$ We shall say that vectors $\mathbf{v}^{(k)}$ tend to the vector $\mathbf{v}$ in the limit if

$$\|\mathbf{v} - \mathbf{v}^{(k)}\| \to 0, \quad \text{i.e.} \quad \|\mathbf{v} - \mathbf{v}^{(k)}\|^2 \to 0. \tag{243}$$

On writing $v_1, v_2, \ldots$ for the components of $\mathbf{v}$, we can express condition (243) in the unabbreviated form

$$\lim_{k\to\infty} \left[\, |\mathbf{v}_1 - \mathbf{v}_1^{(k)}|^2 + |v_2 - v_2^{(k)}|^2 + \ldots \right] = 0. \tag{244}$$

A sum of positive terms tending to zero implies that each term tends to zero, i.e. we have directly from (244):

$$|v_m - v_m^{(k)}| \to 0 \quad \text{as} \quad k \to \infty \ (m = 1,\ 2, \ldots), \tag{245}$$

so that each component $v_m^{(k)}$ must tend to the corresponding component v_m, or more precisely, the real and imaginary parts of $v_m^{(k)}$ must tend

to the real and imaginary parts of v_m. We notice that the converse is not valid, i.e. condition (244) does not necessarily follow from (245). Suppose for instance that $\mathbf{v}^{(k)}$ has the components $(0, \ldots, 0, 1, 0, \ldots)$, the unity being in the kth place. Each component becomes zero on indefinite increase of k, i.e. we have $v_m^{(k)} \to 0$ for any integral m, i.e. $v_m = 0$ $(m = 1, 2, \ldots)$, whereas the sum (244) remains throughout equal to unity.

If the $\mathbf{v}^{(k)}$ sequence tends to $\mathbf{v}$, we write $\mathbf{v}^{(k)} \Rightarrow \mathbf{v}$. We consider the following example of convergence. Let $v(v_1, v_2, \ldots)$ be a given vector and let vectors $\mathbf{v}^{(k)}$ be defined so that their first k components are the same as those of $\mathbf{v}$ whilst the remaining components are zero, i.e.:

$$\mathbf{v}^{(k)}(v_1,\ v_2, \ldots,\ v_k,\ 0,\ 0, \ldots).$$

It is easily shown that $\mathbf{v}^{(k)} \Rightarrow \mathbf{v}$. For in the present case

$$\|\mathbf{v} - \mathbf{v}^{(k)}\|^2 = \sum_{n=k+1}^{\infty} |v_n|^2,$$

and in view of the convergence of the series with the general term $|v_m|^2$, the sum above tends to zero on indefinite increase of k. We observe some simple rules relating to limits. If $\mathbf{u}^{(k)} \Rightarrow \mathbf{u}$ and $\mathbf{v}^{(k)} \Rightarrow \mathbf{v}$, we have

$$\mathbf{u}^{(k)} + \mathbf{v}^{(k)} \Rightarrow \mathbf{u} + \mathbf{v} \quad \text{and} \quad (\mathbf{u}^{(k)}, \mathbf{v}^{(k)}) \to (\mathbf{u}, \mathbf{v}).$$

It may be mentioned that the scalar product is a complex number, which is why we write $\to$ instead of $\Rightarrow$ in the last expression. This last expression in fact shows the continuity of a scalar product. We have by (239):

$$\|(\mathbf{u} + \mathbf{v}) - (\mathbf{u}^{(k)} + \mathbf{v}^{(k)})\| = \|(\mathbf{u} - \mathbf{u}^{(k)}) + (\mathbf{v} - \mathbf{v}^{(k)})\| \leqslant$$
$$\leqslant \|\mathbf{u} - \mathbf{u}^{(k)}\| + \|\mathbf{v} - \mathbf{v}^{(k)}\|,$$

whilst by the definition of limit, $\|\mathbf{u} - \mathbf{u}^{(k)}\| \to 0$ and $\|\mathbf{v} - \mathbf{v}^{(k)}\| \to 0$. It follows from the inequality that

$$\|(\mathbf{u} + \mathbf{v}) - (\mathbf{u}^{(k)} + \mathbf{v}^{(k)})\| \to 0,$$

i.e. in fact $\mathbf{u}^{(k)} + \mathbf{v}^{(k)} \Rightarrow \mathbf{u} + \mathbf{v}$. Furthermore, it follows from the definition of limit that

$$\mathbf{u}^{(k)} = \mathbf{u} + \mathbf{s}^{(k)};\ \mathbf{v}^{(k)} = \mathbf{v} + \mathbf{t}^{(k)},$$

where $\|\mathbf{s}^{(k)}\| \to 0$ and $\|\mathbf{t}^{(k)}\| \to 0$. We have for the scalar product:

$$(\mathbf{u}^{(k)}, \mathbf{v}^{(k)}) = (\mathbf{u} + \mathbf{s}^{(k)}, \mathbf{v} + \mathbf{t}^{(k)}) = (\mathbf{u}, \mathbf{v}) + (\mathbf{u}, \mathbf{t}^{(k)}) + (\mathbf{s}^{(k)}, \mathbf{v}) + (\mathbf{s}^{(k)}, \mathbf{t}^{(k)}).$$

whence

$$|(\mathbf{u}, \mathbf{v}) - (\mathbf{u}^{(k)}, \mathbf{v}^{(k)})| \leqslant |(\mathbf{u}, \mathbf{t}^{(k)})| + |(\mathbf{s}^{(k)}, \mathbf{v})| + |(\mathbf{s}^{(k)}, \mathbf{t}^{(k)})|,$$

or, by (238):

$$|(\mathbf{u}, \mathbf{v}) - (\mathbf{u}^{(k)}, \mathbf{v}^{(k)})| \leqslant \|\mathbf{u}\| \cdot \|\mathbf{t}^{(k)}\| + \|\mathbf{s}^{(k)}\| \cdot \|\mathbf{v}\| + \mathbf{s}^{(k)}\| . \|\mathbf{t}^{(k)}\|.$$

The right-hand side tends to zero, so that also

$$|(\mathbf{u}, \mathbf{v}) - (\mathbf{u}^{(k)}, \mathbf{v}^{(k)})| \to 0 \quad \text{i.e.} \quad (\mathbf{u}^{(k)}, \mathbf{v}^{(k)}) \to (\mathbf{u}, \mathbf{v}).$$

In particular, $(\mathbf{u}^{(k)}, \mathbf{u}^{(k)}) \to (\mathbf{u}, \mathbf{u})$, i.e. $\|\mathbf{u}^{(k)}\|^2 \to \|\mathbf{u}\|^2$ or $\|\mathbf{u}^{(k)}\| \to$ $\to \|\mathbf{u}\|$.

It is easily shown also that if the numerical sequence c_k has the limit c, we have $c_k\mathbf{u}^{(k)} \Rightarrow c\mathbf{u}$.

The necessary and sufficient condition for the existence of a limit is expressed as usual by Cauchy's test. We shall state the test for a given case. Suppose we have the vector sequence

$$\mathbf{v}^{(k)} \quad (k = 1, 2, \ldots). \tag{246}$$

The necessary and sufficient condition for this sequence to have a limit is as follows: given any small positive ε, there exists an N such that

$$\|\mathbf{v}^{(n)} - \mathbf{v}^{(m)}\| < \varepsilon, \tag{247}$$

provided only that n and m are $> N$.

We first show that the condition is necessary. Let sequence (246) have the limit $\mathbf{v}$. We can now write

$$\mathbf{v}^{(n)} - \mathbf{v}^{(m)} = (\mathbf{v}^{(n)} - \mathbf{v}) + (\mathbf{v} - \mathbf{v}^{(m)}),$$

and therefore, by the triangle rule:

$$\|\mathbf{v}^{(n)} - \mathbf{v}^{(m)}\| \leqslant \|\mathbf{v}^{(n)} - \mathbf{v}\| + \|\mathbf{v} - \mathbf{v}^{(m)}\|.$$

It follows at once from the definition of limit that both terms on the right tend to zero on increase of n and m, so that the same must be true for the term on the left, i.e. condition (247) must in fact be satisfied. We now turn to the sufficiency of (247). We assume that (247) is fulfilled, and show that the sequency (246) tends to a limit. We can write (247) in the expanded form:

$$\sum_{s=1}^{\infty} |v_s^{(n)} - v_s^{(m)}|^2 < \varepsilon^2 \quad \text{for } n \text{ and } m > N, \tag{248}$$

the components of $\mathbf{v}^{(j)}$ being written $v_s^{(j)}$. It follows at once that for any s:

$$|v_s^{(n)} - v_s^{(m)}| < \varepsilon \quad \text{for} \quad n \quad \text{and} \quad m > N$$

or alternatively, on separating into the real and imaginary parts:

$$v_s^{(n)} = \alpha_s^{(n)} + i\beta_s^{(n)},$$

we can write

$$|\alpha_s^{(n)} - \alpha_s^{(m)}| < \varepsilon \quad \text{and} \quad |\beta_s^{(n)} - \beta_s^{(m)}| < \varepsilon.$$

We can say by applying the ordinary Cauchy test that $\alpha_s^{(n)}$ and $\beta_s^{(n)}$ have the limits α_s and β_s, and consequently $v_s^{(n)}$ tends to the complex number $\alpha_s + i\,\beta_s$. We call this limit v_s and show first that the series $\sum\limits_{s=1}^{\infty} |v_s|^2$ is convergent, i.e. the v_s are the components of a vector. On retaining a finite number of first terms in sum (248) and passing to the limit as $n \to \infty$ in this finite sum, we get

$$\sum_{s=1}^{M} |v_s - v_s^{(m)}|^2 \leqslant \varepsilon^2,$$

where M is any integer. Passage to the limit with $M \to \infty$ in this last expression gives us

$$\sum_{s=1}^{\infty} |v_s - v_s^{(m)}|^2 \leqslant \varepsilon^2, \tag{249}$$

whence it follows at once that the numbers $v_s - v_s^{(m)}$ are the components of a vector. We already know that this is true as regards the numbers $v_s^{(m)}$, and we can therefore say that it is true for the v_s, i.e. the v_s are the components of a vector $\mathbf{v}$. We can thus write (249) as

$$\|\mathbf{v} - \mathbf{v}^{(m)}\| < \varepsilon,$$

for $m > N$, i.e. $\mathbf{v}^{(m)} \Rightarrow \mathbf{v}$, and sequence (246) in fact has a limit. Each component v_s of the vector $\mathbf{v}$ is obviously defined as the limit of $v_s^{(m)}$, whence it follows at once that there can only be the one limit. We now consider the infinite vector sum

$$\mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \dots \tag{250}$$

It is said to be convergent if the sum of the first n terms:

$$\mathbf{s}^{(n)} = \mathbf{u}^{(1)} + \dots + \mathbf{u}^{(n)}$$

has a limit in the above sense as $n \to \infty$. By Cauchy's test, the necessary and sufficient condition for convergence is that

$$\|\mathbf{s}^{(n+p)} - \mathbf{s}^{(n)}\| = \|\mathbf{u}^{(n+1)} + \dots + \mathbf{u}^{(n+p)}\| \leqslant \varepsilon, \tag{251}$$

for $n > N$ and any p.

We have, on taking into account the continuity of scalar products:

$$(\mathbf{x}, \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \ldots) = (\mathbf{x}, \mathbf{u}^{(1)}) + (\mathbf{x}, \mathbf{u}^{(2)}) + \ldots$$
$$(\mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \ldots, \mathbf{x}) = (\mathbf{u}^{(1)}, \mathbf{x}) + (\mathbf{u}^{(2)}, \mathbf{x}) + \ldots$$

On applying this to the case when the vectors $\mathbf{u}^{(k)}$ are mutually orthogonal, we have

$$(\mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \ldots, \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \ldots) = (\mathbf{u}^{(1)}, \mathbf{u}^{(1)}) + (\mathbf{u}^{(2)}, \mathbf{u}^{(2)}) + \ldots$$

or

$$\| \mathbf{u}^{(1)} + \mathbf{u}^{(2)} + \ldots \|^2 = \| \mathbf{u}^{(1)} \|^2 + \| \mathbf{u}^{(2)} \|^2 + \ldots,$$

i.e. Pythagoras' theorem is also valid for the sum of an infinite set of mutually orthogonal vectors.

We now establish the necessary and sufficient condition for the convergence of series (250) of mutually orthogonal vectors. In accordance with Cauchy's test, we have to form expression (251) which is equal, by Pythagoras' theorem, to

$$\| \mathbf{u}^{(n+1)} \|^2 + \ldots + \| \mathbf{u}^{(n+p)} \|^2 .$$

Hence it follows at once that the necessary and sufficient condition for convergence of the series is the convergence of the series consisting of the squares of the norms of the vectors $\mathbf{u}^{(k)}$. This result can be expressed alternatively as follows: let $\mathbf{x}^{(k)}$ be mutually orthogonal unit vectors. We form the series

$$\sum_{k=1}^{\infty} C_k \mathbf{x}^{(k)}, \tag{252}$$

where the C_k are certain numbers. The necessary and sufficient condition for the convergence of this series is, by what we have proved above, the convergence of the series

$$\sum_{k=1}^{\infty} | C_k |^2 .$$

This implies among other things that changing the order of the terms in series (252) does not affect its convergence. It is easy to show also that the sum of (252) remains unchanged on changing the order of the terms.

47. Complete systems of mutually orthogonal vectors. We now bring in an important concept, that *of a complete system of mutually orthogonal vectors*. We can show, as in the case of a finite number of dimen-

sions, that every finite set of mutually orthogonal vectors is linearly independent. We saw in the case of n-dimensional space that a set of any n linearly independent vectors formed a complete system in the sense that any vector could be expressed as a linear combination of these n vectors. We no longer have such a simple criterion of completeness in the case of an H space, since the number of dimensions is infinite. We shall only employ mutually orthogonal unit vectors in future.

Suppose we have an infinite set of mutually orthogonal unit vectors $\mathbf{x}^{(k)}$ $(k = 1, 2, \ldots)$, and let $\mathbf{y}$ be a given vector of the space H. As in the case of a finite number of vectors, we form the sum of the projections of our vector on the axes:

$$\sum_{k=1}^{\infty} (\mathbf{y}, \mathbf{x}^{(k)})\, \mathbf{x}^{(k)}. \tag{253}$$

As shown above, we have the inequality for any m:

$$\sum_{k=1}^{m} |(\mathbf{y}, \mathbf{x}^{(k)})|^2 \leqslant \|\mathbf{y}\|^2,$$

and therefore in the limit

$$\sum_{k=1}^{\infty} |(\mathbf{y}, \mathbf{x}^{(k)})|^2 \leqslant \|\mathbf{y}\|^2, \tag{254}$$

so that the series on the left must be convergent. It now follows at once that series (253) must also be convergent. Suppose

$$\mathbf{y} = \sum_{k=1}^{\infty} (\mathbf{y}, \mathbf{x}^{(k)})\, \mathbf{x}^{(k)} + \mathbf{u}. \tag{255}$$

It may readily be shown as in [45] that the vector $\mathbf{u}$ is orthogonal to all the vectors $\mathbf{x}^{(k)}$, and consequently, by Pythagoras' theorem:

$$\|\mathbf{y}\|^2 = \sum_{k=1}^{\infty} |(\mathbf{y}, \mathbf{x}^{(k)})|^2 + \|\mathbf{u}\|^2. \tag{256}$$

Hence it follows that if the vector $\mathbf{u}$ in (255) differs from zero, we have the $<$ sign in (254), whilst if $\mathbf{u}$ is zero (i.e. all its components vanish), we have the $=$ sign in (254).

The system of axes $\mathbf{u}^{(k)}$ is said to be *complete* if we have the $=$ sign in (254) for any vector $\mathbf{y}$ of the H space. In this case, we can evidently

resolve any vector in terms of the complete system of fundamental vectors:

$$\mathbf{y} = \sum_{k=1}^{\infty} (\mathbf{x}^{(k)} \cdot \mathbf{y})\, \mathbf{x}^{(k)}. \tag{257}$$

A complete system is alternatively said to be *closed*, whilst the formula

$$\sum_{k=1}^{\infty} |(\mathbf{y}, \mathbf{x}^{(k)})|^2 = \|\mathbf{y}\|^2 \tag{258}$$

is called the *closure equation*. We notice a consequence of (258), called the *generalized closure equation*. Let us have two vectors $\mathbf{y}$ and $\mathbf{z}$, and let the $\mathbf{x}^{(k)}$ form a complete system, so that for any $\mathbf{y}$ and $\mathbf{x}$:

$$\mathbf{y} = \sum_{k=1}^{\infty} (\dot{\mathbf{y}}, \mathbf{x}^{(k)})\, \mathbf{x}^{(k)}; \quad \mathbf{z} = \sum_{k=1}^{\infty} (\mathbf{z}, \mathbf{x}^{(k)})\, \mathbf{x}^{(k)}. \tag{259}$$

On applying (258) to the vectors $\mathbf{y} + \mathbf{z}$ and $\mathbf{y} + i\mathbf{z}$, we get:

$$\sum_{k=1}^{\infty} [(\mathbf{y}, \mathbf{x}^{(k)}) + (\mathbf{z}, \mathbf{x}^{(k)})]\,[\overline{(\mathbf{y}, \mathbf{x}^{(k)})}] + \overline{(\mathbf{z}, \mathbf{x}^{(k)})}] = (\mathbf{y} + \mathbf{z}, \mathbf{y} + \mathbf{z}),$$

$$\sum_{k=1}^{\infty} [(\mathbf{y}, \mathbf{z}^{(k)}) + i(\mathbf{z}, \mathbf{x}^{(k)})]\;[\overline{(\mathbf{y}, \mathbf{x}^{(k)})} - i\,\overline{(\mathbf{z}, \mathbf{x}^{(k)})}] = (\mathbf{y} + i\mathbf{z}, \mathbf{y} + i\mathbf{z}).$$

We obtain, on using the closure equation for $\mathbf{y}$ and $\mathbf{z}$:

$$\sum_{k=1}^{\infty} (\mathbf{y}, \mathbf{x}^{(k)})\,\overline{(\mathbf{z}, \mathbf{x}^{(k)})} + \sum_{k=1}^{\infty} (\mathbf{z}, \mathbf{x}^{(k)})\,\overline{(\mathbf{y}, \mathbf{x}^{(k)})} = (\mathbf{y}, \mathbf{z}) + (\mathbf{z}, \mathbf{y}),$$

$$\sum_{k=1}^{\infty} (\mathbf{y}, \mathbf{x}^{(k)})\,\overline{(\mathbf{z}, \mathbf{x}^{(k)})} - \sum_{k=1}^{\infty} (\mathbf{z}, \mathbf{x}^{(k)})\,\overline{(\mathbf{y}, \mathbf{x}^{(k)})} = (\mathbf{y}, \mathbf{z}) - (\mathbf{z}, \mathbf{y}),$$

whence follows the generalized closure equation:

$$\sum_{k=1}^{\infty} (\mathbf{y}, \mathbf{x}^{(k)})\,\overline{(\mathbf{y}, \mathbf{x}^{(k)})} = (\mathbf{y}, \mathbf{z}). \tag{260}$$

If $\mathbf{x}$ is the same as $\mathbf{y}$, this equation becomes (258).

We now consider in detail the fundamental vectors $\mathbf{x}^{(k)}$. Since these are mutually orthogonal unit vectors, we have for their components $\mathbf{x}_s^{(k)}$ $(s = 1, 2, \ldots)$:

$$\sum_{s=1}^{\infty} x_s^{(p)}\,\overline{x}_s^{(q)} = \delta_{pq}, \tag{261}$$

where $\delta_{pq} = 0$ for $p \neq q$ and $\delta_{pp} = 1$.

We now find the condition for completeness of the system $\mathbf{x}^{(k)}$. We bring in for this purpose the vectors $\mathbf{y}^{(l)}$, whose lth components are equal to unity and the rest zero. We have

$$(\mathbf{y}^{(l)}, \mathbf{x}^{(k)}) = x_l^{(k)},$$

and (258) gives

$$\sum_{k=1}^{\infty} |x_l^{(k)}|^2 = 1 \qquad (l = 1, 2, 3, \ldots).$$

On now applying (260) to the vectors $\mathbf{y}^{(p)}$ and $\mathbf{y}^{(q)}$ for $p \neq q$, and using the fact that they are orthogonal, we obtain in addition to the above the following conditions:

$$\sum_{k=1}^{\infty} x_p^{(k)} x_q^{(k)} = 0 \; (p \neq q),$$

i.e. in general,

$$\sum_{k=1}^{\infty} x_p^{(k)} \bar{x}_q^{(k)} = \delta_{pq}. \tag{262}$$

We write down the components of our vectors $\mathbf{x}^{(k)}$ in the form of an infinite matrix:

$$\begin{array}{l} x_1^{(1)}, x_1^{(2)}, x_1^{(3)}, \ldots \\ x_2^{(1)}, x_2^{(2)}, x_2^{(3)}, \ldots \\ \cdots\cdots\cdots\cdots \end{array} \tag{263}$$

Equations (261), expressing the fact that the $\mathbf{x}^{(k)}$ are mutually orthogonal unit vectors, are equivalent to the fact that the columns of this matrix are normalized and orthogonal. Conditions (262) show that the rows must also be normalized and orthogonal for the system of $\mathbf{x}^{(k)}$ to be complete.

We now show that conditions (262) with $p = q$ are likewise sufficient for completeness. In fact, if these conditions are satisfied with $p = q$, the closure equation applies for the vectors

$$\mathbf{y}^{(l)} (0, \ldots, 0, \overset{l}{1}, 0, \ldots)$$

and all these can be expressed linearly in terms of the vectors $\mathbf{x}^{(k)}$:

$$\mathbf{y}^{(l)} = \sum_{k=1}^{\infty} c_k^{(l)} \mathbf{x}^{(k)}.$$

We show that the same is true for any vector $\mathbf{z}$. We denote by $\mathbf{z}^{(l)}$ the vector whose first l components are the same as for $\mathbf{z}$, whilst the

remaining components are zero. We obviously have

$$\mathbf{z}^{(l)} = z_1 \mathbf{y}^{(1)} + \ldots + z_1 \mathbf{y}^{(l)},$$

and since the $\mathbf{y}^{(m)}$ are expressible linearly in terms of the $\mathbf{x}^{(k)}$, the same can be said of $\mathbf{z}^{(l)}$:

$$\mathbf{z}^{(l)} = \sum_{k=1}^{\infty} d_k^{(l)} \mathbf{x}^{(k)}.$$

On forming the scalar product of both sides with $\mathbf{x}^{(k)}$, we get the following usual expressions for the coefficients $d_k^{(l)}$:

$$d_k^{(l)} = (\mathbf{z}^{(l)}, \mathbf{x}^{(k)}).$$

On the other hand, as we have seen above:

$$\mathbf{z} = \sum_{k=1}^{\infty} d_k \mathbf{x}^{(k)} + \mathbf{u}, \tag{264}$$

where $\mathbf{u}$ is orthogonal to all the $\mathbf{x}^{(k)}$. We now consider the difference

$$z - z^{(l)} = \sum_{k=1}^{\infty} (d_k - d_k^{(l)}) \mathbf{x}^{(k)} + \mathbf{u}.$$

By Pythagoras' theorem:

$$\| z - z^{(l)} \|^2 = \sum_{k=1}^{\infty} \| d_k - d_k^{(l)} \|^2 + \| \mathbf{u} \|^2,$$

and consequently

$$\| \mathbf{u} \|^2 \leqslant \| z - z^{(l)} \|^2.$$

The vector $\mathbf{u}$ does not depend on l, whilst we know from [46] that the right-hand side tends to zero as $l \to \infty$. Hence it follows at once that $\mathbf{u} = 0$, and (264) gives the resolution of any vector $\mathbf{z}$ in terms of the $\mathbf{x}^{(k)}$:

$$\mathbf{z} = \sum_{k=1}^{\infty} d_k \mathbf{x}^{(k)} \quad [d_k = \mathbf{x}^{(k)} \cdot \mathbf{z}]. \tag{265}$$

Thus the closure equation is valid for any vector. The final result can be stated as follows. *The necessary and sufficient condition for mutually orthogonal unit vectors* $\mathbf{x}^{(k)}$ *to form a complete (closed) system is that the sum of squares of the moduli of the elements of each row of matrix* (263) *is equal to unity.* If this condition is satisfied for matrix (263), its rows are orthogonal.

48. Linear transformations with an infinite set of variables. We shall consider in brief outline the linear transformation with an infinite set of variables:

$$\left.\begin{aligned} x_1' &= a_{11}x_1 + a_{12}x_2 + \cdots \\ x_2' &= a_{21}x_1 + a_{22}x_2 + \cdots \\ &\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{266}$$

or

$$\mathbf{x}' = A\mathbf{x}, \tag{267}$$

where A is the infinite matrix with elements a_{ik}. We first of all lay down the condition that the infinite series on the right-hand sides of equations (266) are convergent for any vector $\mathbf{x}$ of the space H. As we know, this condition is satisfied if the series

$$\sum_{k=1}^{\infty} |a_{ik}|^2 \qquad (i = 1, 2, \ldots)$$

are convergent for any i. It can be shown that this condition is necessary as well as sufficient. If this condition is not satisfied, the series on the right-hand sides of equations (266) are convergent for only a part, and not the whole, of the space H.

It is natural to lay down the further condition that if x_k is a vector component, the number x_k' obtained as a result of transformation (266) also represents a vector component of the space H, i.e. the series

$$\sum_{k=1}^{\infty} |x_k'|^2$$

must be convergent if we have convergence of the series

$$\sum_{k=1}^{\infty} |x_k|^2.$$

If the matrix A satisfies the above two conditions, the corresponding transformation A is said to be *bounded*. The point of this term lies in the fact that we can prove the existence for such a transformation of a positive number M such that

$$\|\mathbf{x}'\|^2 \leqslant M\|\mathbf{x}\|^2, \tag{268}$$

or in the expanded form:

$$\sum_{k=1}^{\infty} |x_k'|^2 \leqslant M \sum_{k=1}^{\infty} |x_k|^2. \tag{269}$$

We shall dwell on a particular case of a linear transformation. We take the transformation

$$\left.\begin{aligned} x_1' &= u_{11}x_1 + u_{12}x_2 + \cdots \\ x_2' &= u_{21}x_1 + u_{22}x_2 + \cdots \\ &\cdots\cdots\cdots\cdots \end{aligned}\right\}, \tag{270}$$

the series

$$\sum_{k=1}^{\infty} |u_{ik}|^2$$

being as usual assumed convergent for any i. We bring into the discussion vectors $\mathbf{u}^{(k)}$ with components $\bar{u}_{k1}, \bar{u}_{k2}, \ldots$, and suppose that the coefficients u_{ik} are such that the vectors $\mathbf{u}^{(k)}$ form a complete system of mutually orthogonal unit vectors. As we have shown above, this is equivalent to the rows and columns of the matrix of u_{ik} being orthogonal and normalized, i.e.

$$\sum_{s=1}^{\infty} u_{sp}\overline{u_{sq}} = \delta_{pq}, \qquad \sum_{s=1}^{\infty} u_{ps}\overline{u_{qs}} = \delta_{pq}. \tag{271}$$

The corresponding transformation (270) is said to be *unitary* in this case.

We can write equations (270) as

$$\left.\begin{aligned} (\mathbf{x}, \mathbf{u}^{(1)}) &= x_1' \\ (\mathbf{x}, \mathbf{u}^{(2)}) &= x_2' \\ &\cdots\cdots \end{aligned}\right\} \tag{272}$$

The closure formula gives us

$$\sum_{k=1}^{\infty} |x_k'|^2 = \|\mathbf{x}\|^2 = \sum_{k=1}^{\infty} |x_k|^2,$$

i.e. as in the case of a finite number of dimensions, a unitary transformation does not change the length of a vector, and we can take $M = 1$ in expression (268).

Equations (270) may readily be solved with respect to the $\mathbf{x}_k$, which gives us the inverse transformation to (270). On using the fact that the system $\mathbf{u}^{(k)}$ is complete, we obtain from equations (272) the following definitive expression for the vector $\mathbf{x}$:

$$\mathbf{x} = x_1'\mathbf{u}^{(1)} + x_2'\mathbf{u}^{(2)} + \cdots \tag{273}$$

or

$$\left.\begin{aligned} x_1 &= \bar{u}_{11}x_1' + \bar{u}_{21}x_2' + \cdots \\ x_2 &= \bar{u}_{12}x_2' + \bar{u}_{22}x_2' + \cdots \\ &\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{274}$$

In other words, if equations (270) have a solution, it must be expressed by (273) or (274). Of course, we are referring here only to the solutions x_k for which the sums of the squares of the moduli are convergent. We now show that equation (273) in fact yields the solution of the problem. The given numbers x'_k are by hypothesis such that the squares of the moduli form a convergent series. Hence follows the convergence of series (273), as we know, since the $\mathbf{u}^{(k)}$ are mutually orthogonal unit vectors. We have for the sum of this series;

$$(\mathbf{x}, \mathbf{u}^{(k)}) = (x'_1 \mathbf{u}^{(1)} + x'_2 \mathbf{u}^{(2)} + \ldots, \mathbf{u}^{(k)}) = x'_k,$$

i.e. the sum in fact satisfies system (270). System (274) shows that the inverse of the unitary transformation is obtained by replacing rows by columns and all the elements by their conjugates, i.e. we have here an entirely analogous case to that of a finite number of dimensions.

In the general case even of bounded matrices, the problems of inverse matrices and of reduction to the diagonal form present greater difficulty and lead to results that have no strict analogue in finite-dimensional space. A more detailed account of linear transformations by means of infinite matrices will be found in the fifth volume. We confine ourselves here to indicating a few results. We may mention the necessary and sufficient condition as regards the coefficients a_{ik} for transformation (266) to be bounded. It is stated thus: there exists a positive number N such that, with any positive integral k and any numbers x_s $(s = 1, 2, \ldots)$, we have the inequality

$$\left| \sum_{n, m=1}^{k} a_{nm} x_m \bar{x}_n \right| \leqslant N \sum_{m=1}^{k} |x_m|^2 .$$

The following simple sufficient condition may also be proved for boundedness of the transformation (266): there exists a positive number l (not dependent on m or n) such that we have the inequalities

$$\sum_{m=1}^{\infty} |a_{nm}| \leqslant l; \quad \sum_{n=1}^{\infty} |a_{nm}| \leqslant l .$$
$$(n=1, 2, \ldots) \qquad (m=1, 2, \ldots)$$

If the matrix A defines a bounded transformation, there exists a unique matrix $\tilde{A}$ such that, for any $\mathbf{x}$ and $\mathbf{y}$:

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \tilde{A}\mathbf{y}),$$

the elements $\tilde{a}_{ik}$ of $\tilde{A}$ being given by $\tilde{a}_{ik} = \bar{a}_{ki}$. If $\tilde{A}$ is the same as A, i.e. $a_{ik} = \bar{a}_{ki}$, the bounded transformation (266) is called Hermitian or self-conjugate.

We have for bounded transformations:

$$(A\mathbf{x}, \mathbf{y}) = \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{nm} x_m\right) \bar{y}_n = \sum_{m=1}^{\infty} x_m \left(\sum_{n=1}^{\infty} a_{nm} \bar{y}_n\right) =$$

$$= \lim_{\substack{k\to\infty \\ l\to\infty}} \sum_{m=1}^{k} \sum_{n=1}^{l} a_{nm} x_m \bar{y}_n .$$

We notice the important particular case of a bounded operator, when we have convergence of the double series

$$\sum_{n,m=1}^{\infty} |a_{nm}|^2. \tag{275}$$

In this case the double series

$$\sum_{n,m=1}^{\infty} a_{nm} x_m \bar{y}_n$$

is absolutely convergent for any choice of vectors $\mathbf{x}(x_1, x_2, \ldots)$ and $\mathbf{y}(y_1, y_2, \ldots)$. If, in addition to the convergence of series (275), we have $a_{ik} = \bar{a}_{ki}$, we arrive at the possibility of reducing the Hermitian form to a sum of squares with the aid of a unitary transformation:

$$\sum_{n,m=1}^{\infty} a_{nm} x_m \bar{y}_n = \sum_{k=1}^{\infty} \lambda_k z_k \bar{z}_k ,$$

where the vector $\mathbf{z}(z_1, z_2, \ldots)$ is obtained by application of a unitary transformation to the vector $\mathbf{x}(x_1, \mathbf{x}_2, \ldots)$: $\mathbf{z} = U\mathbf{x}$. With this, $\lambda_k \to 0$ as $k \to \infty$. If A and B are infinite matrices, yielding bounded transformations, successive application of them also yields a bounded transformation, the coefficients of which are given by the usual expressions

$$\{BA\}_{ik} = \sum_{s=1}^{\infty} \{B\}_{is} \{A\}_{sk} .$$

We remark also that, if the vector sequence $\mathbf{x}^{(k)}$ has the limit $\mathbf{x}$, i.e. $\mathbf{x}^{(k)} \Rightarrow \mathbf{x}$, then $A\mathbf{x}^{(k)} \Rightarrow A\mathbf{x}$, if A is the matrix of a bounded transformation.

Unbounded linear transformations also play an essential part in applications to mathematical physics. These are discussed in the fifth volume.

49. Functional space. We have considered the space H in which a vector is defined by an infinite set of components, enumerated by means of integers: the first component being x_1, the second x_2, and

so on. We now turn to the functional space F in which *the role of a vector is played by a function* of one or more arguments which are capable of continuous variation.

We consider a function $f(x)$, defined in the interval $a \leqslant x \leqslant b$. We can regard the function as a vector; for every value x_0 of the above interval there is a corresponding number $f(x_0)$ which gives the component of the vector with the subscript x_0. Here the independent variable x, which plays the part of component subscript, runs continuously through all the values in the interval $a \leqslant x \leqslant b$, so that our vector $f(x)$ has a continuous set of components. The value x_0 corresponds to the number of an axis in previous notations, whilst the value of the function $f(x_0)$ gives the magnitude of the corresponding component. We shall assume here that $f(x)$ can take both real and complex values, whilst the interval of variation of the independent variable will always be taken to be a finite segment of the real axis.

For the present we shall consider for the sake of definiteness the complex functions $f(x) = f_1(x) + if_2(x)$, defined and continuous in the finite interval $a \leqslant x \leqslant b$.

Such functions can be multiplied by complex numbers and added, as in the case of vectors of space H. This leads to further continuous functions. When defining the norm and scalar product we must replace summations everywhere by integrations. A scalar product is defined by

$$(\varphi(x), \psi(x)) = \int_a^b \varphi(x)\, \overline{\psi(x)}\, dx \tag{276}$$

and the square of the norm by

$$\| f(x) \|^2 = (f(x), f(x)) = \int_a^b | f(x) |^2\, dx\,. \tag{277}$$

Let the system of functions $\varphi_k(x)$ $(k = 1, 2, \ldots)$ form a system of mutually orthogonal unit vectors, i.e.

$$\int_a^b \varphi_p(x)\, \overline{\varphi_q(x)}\, dx = \delta_{pq}\,. \tag{278}$$

We have already mentioned such systems of normalized and orthogonal functions [II, 148] and we confine ourselves here to recalling some results that have a direct connection with the above. The only new feature compared with [II, 148] is the fact that our present functions can also take complex values.

Suppose, then, that the $\varphi_k(x)$ form an orthogonal normalized system and let $f(x)$ be a given vector (or function). We bring into the discussion the Fourier coefficients of $f(x)$ or, in our present terminology, the magnitudes of the projections of the vector $f(x)$ on the axes of the functional space represented by the functions $\varphi_k(x)$:

$$a_k = (f(x), \varphi_k(x)) = \int_a^b f(x)\, \overline{\varphi_k(x)}\, \mathrm{d}x\,. \tag{279}$$

We consider the integral

$$I_n = \int_a^b \Big| f(x) - \sum_{k=1}^{n} a_k\, \varphi_k(x) \Big|^2 \mathrm{d}x \tag{280}$$

or

$$I_n = \int_a^b \Big[f(x) - \sum_{k=1}^{n} a_k\, \varphi_k(x) \Big] \Big[\overline{f(x)} - \sum_{k=1}^{n} \overline{a_k\, \varphi_k(x)} \Big] \mathrm{d}x\,.$$

We take into account equations (278) and (279) and arrive at the following expression for the integral:

$$I_n = \int_a^b |f(x)|^2\, \mathrm{d}x - \sum_{k=1}^{n} |a_k|^2,$$

whilst in view of the fact that $I_n \geqslant 0$, we have

$$\sum_{k=1}^{n} |a_k|^2 \leqslant \int_a^b |f(x)|^2\, \mathrm{d}x \tag{281}$$

and in the limit, as $n \to \infty$:

$$\sum_{k=1}^{\infty} |a_k|^2 \leqslant \int_a^b |f(x)|^2\, \mathrm{d}x\,. \tag{282}$$

This is known as *Bessel's inequality.*

If we have the $=$ sign in the last expression (282), the integral I_n tends to zero on indefinite increase of n and, conversely, if the integral tends to zero, we have the $=$ sign in (282).

If the $=$ sign is obtained in (282), i.e.

$$\sum_{k=1}^{\infty} |a_k|^2 = \int_a^b |f(x)|^2\, \mathrm{d}x \tag{283}$$

for any continuous function $f(x)$, the system of functions $\varphi_k(x)$ is said to be *complete* or *closed,* whilst equation (283) is called *the closure equation.*

The integral I_n here tends to zero for any continuous function $f(x)$:

$$\lim_{n\to\infty} \int_a^b \Big| f(x) - \sum_{k=1}^{n} a_k\, \varphi_k(x) \Big| \,\mathrm{d}x = 0\,, \tag{284}$$

i.e. any such function can be represented to any required degree of accuracy by a linear combination of a finite number of $\varphi_k(x)$, the phrase "to any required degree of accuracy" being understood to mean, not arbitrary smallness of the difference itself:

$$\Big| f(x) - \sum_{k=1}^{n} a_k\, \varphi_k(x) \Big|\,,$$

but arbitrary smallness of the integral I_n for large n. Thus if we want to be more precise, we should speak of approximating to $f(x)$ by a linear combination of a finite number of $\varphi_k(x)$ to as small as desired a root mean square error.

Just as in [47], a *generalized closure equation* may be written for a complete system of functions $\varphi_k(x)$. In fact, let a_k and b_k be the Fourier coefficients of the functions $f(x)$ and $f_1(x)$:

$$a_k = \int_a^b f(x)\,\overline{\varphi_k(x)}\,\mathrm{d}x; \quad b_k = \int_a^b f_1(x)\,\overline{\varphi_k(x)}\,\mathrm{d}x\,. \tag{285}$$

The following generalized closure equation is valid:

$$\sum_{k=1}^{\infty} a_k \overline{b}_k = \int_a^b f(x)\,\overline{f_1(x)}\,\mathrm{d}x\,. \tag{286}$$

Let the a_k be the Fourier coefficients of $f(x)$ as above. We form the Fourier series

$$\sum_{k=1}^{\infty} a_k\, \varphi_k(x)\,.$$

We cannot say that this series is convergent, or still less that its sum is equal to $f(x)$. The following notation is commonly used:

$$f(x) \sim \sum_{k=1}^{\infty} a_k\, \varphi_k(x)\,, \tag{287}$$

where the symbol $\sim$ merely indicates that the infinite series on the right is the Fourier series for $f(x)$. Though (287) is not an equation in the ordinary sense of the word, yet as we saw in [II, 148], if the $\varphi_k(x)$

form a complete system, the expression becomes a strict equality on term-by-term integration of the right-hand side, i.e.

$$\int_{x_1}^{x_2} f(x)\,dx = \sum_{k=1}^{\infty} a_k \int_{x_1}^{x_2} \varphi_k(x)\,dx \qquad (a \leqslant x_1 < x_2 \leqslant b)\,.$$

Prior to integration, we can multiply both sides of (287) by the continuous function $\psi(x)$, i.e.

$$\int_{x_1}^{x_2} f(x)\,\psi(x)\,dx = \sum_{k=1}^{\infty} a_k \int_{x_1}^{x_2} \varphi_k(x)\,\psi(x)\,dx\,.$$

Integration over the full interval (a, b) gives us

$$\int_a^b f(x)\,\psi(x)\,dx = \sum_{k=1}^{\infty} a_k \int_a^b \varphi_k(x)\,\psi(x)\,dx\,.$$

It may easily be verified that this expression represents the generalized closure equation for the functions $f(x)$ and $\overline{\psi(x)}$.

50. The connection between functional and Hilbert space. We now undertake to establish the relationship, of great importance in theoretical physics, between the functional space described in the previous section and the space H that we discussed earlier.

Suppose we have the complete system of orthogonal, normalized functions

$$\varphi_k(x) \qquad (k = 1, 2, \ldots) \tag{288}$$

in a functional space, so that equation (283) holds for any continuous function $f(x)$. We take a second continuous function $f_1(x)$ and suppose as above that

$$b_k = \int_a^b f_1(x)\,\overline{\varphi_k(x)}\,dx\,.$$

On applying the closure formula to the difference $f(x) - f_1(x)$, we get

$$\sum_{k=1}^{\infty} |a_k - b_k|^2 = \int_a^b |f(x) - f_1(x)|^2\,dx\,. \tag{289}$$

If the continuous functions $f(x)$ and $f_1(x)$ differ, the right-hand side is certainly greater than zero, and consequently the coefficients b_k cannot all be the same as the a_k, i.e. different continuous functions have different sets of Fourier coefficients with respect to system (288). Hence every continuous function is fully characterized by its Fourier coefficients, the squares of the moduli of which form on summation a

convergent series, i.e. *to every continuous function there corresponds a definite vector of the space H, the vectors corresponding to different continuous functions being themselves different.* Let $f_n(x)$ $(n = 1, 2, \ldots)$ be a sequence of functions with Fourier coefficients $a_k^{(n)}$, i.e.

$$a_k^{(n)} = \int_a^b \overline{\varphi_k(x)}\, f_n(x)\, dx. \tag{290}$$

The closure equation gives

$$\sum_{k=1}^{\infty} |a_k - a_k^{(n)}|^2 = \int_a^b |f(x) - f_n(x)|^2\, dx, \tag{291}$$

whence it follows at once that the convergence of a vector with components $a_k^{(n)}$ $(k = 1, 2, \ldots)$ of the space H to a vector with components a_k is equivalent to the equation in our functional space:

$$\lim_{n\to\infty} \int_a^b |f(x) - f_n(x)|^2\, dx = 0. \tag{292}$$

If we take vectors with components

$$\mathbf{z}\,(a_1, a_2, \ldots) \quad \text{and} \quad \mathbf{z}^{(n)}\,(a_1, \ldots, a_n, 0, 0, \ldots),$$

$\mathbf{z}^{(n)}$ in the functional space corresponds to the part of the Fourier series of $f(x)$:

$$\sum_{k=1}^{n} a_k \varphi_k(x).$$

We know from [45] that $\mathbf{z}^{(n)} \to \mathbf{z}$ which corresponds to the fact that the integral

$$\int_a^b \left| f(x) - \sum_{k=1}^{n} a_k \varphi_k(x) \right|^2 dx$$

tends to zero.

As explained above, to every continuous function of our functional space there corresponds a definite vector of the space H. The converse statement does not hold, i.e. the vectors of the space H corresponding to continuous functions form only part of the space H. If we want the converse to be actually true, we have to consider some wider class of function than that of continuous functions; but this is a problem that we cannot dwell on here.

We have established the correspondence between the functional space of continuous functions and the space H by taking a definite system of orthogonal functions (288) as our starting-point. If we introduce

instead of this the new system of orthogonal and normalized functions

$$\psi_k(x) \qquad (k = 1, 2, \ldots), \tag{293}$$

the law of correspondence is naturally not the same. It can be shown that the vectors of space H corresponding to these latter functions must be subject to a unitary transformation. Of course system (293) must also be complete in this case.

Definite Fourier coefficients with respect to system (288), or in other words, a definite Fourier series, correspond to every function $\psi_m(x)$ of system (293).

We thus have the following array:

$$\psi_m(x) \sim \sum_{k=1}^{\infty} u_{km} \varphi_k(x).$$

The sign $\sim$ merely shows that the function on the left corresponds to the Fourier series on the right. On taking into account the fact that the functions $\psi_k(x)$ are normalized, and the closure formula (283), we have

$$\sum_{k=1}^{\infty} |u_{km}|^2 = 1. \tag{294}$$

In addition, the generalized closure equation holds:

$$\sum_{k=1}^{\infty} \bar{u}_{kp} u_{kp} = \int_a^b \overline{\psi_p(x)} \psi_q(x)\, \mathrm{d}x,$$

which gives us, by the orthogonality of the $\psi_k(x)$ and (294):

$$\sum_{k=1}^{\infty} \bar{u}_{kp} u_{kq} = \delta_{pq}. \tag{295}$$

This expression shows us that the matrix U with elements u_{lk} satisfies the condition for its columns to be normalized and orthogonal. We can show by making use of the results of [48] that the necessary and sufficient condition for the system of functions (293) to be complete is that the sum of the squares of the moduli of the elements of each row of matrix U be equal to unity, i.e.

$$\sum_{k=1}^{\infty} |u_{ik}|^2 = 1 \qquad (i = 1, 2, \ldots). \tag{296}$$

All the above refers to the case when the functional space is made up of functions of a single independent variable. We can also

consider functions of several independent variables, defined in some domain of space of more than one dimension. All our discussion remains in force, the only difference being that the single integrals are everywhere replaced by multiple integrals over the domain in which our functions are defined.

We may take as an example the system of functions

$$\varphi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx} \qquad (k = 0, \pm 1, \pm 2, \ldots) \tag{297}$$

and let the fundamental interval be $(-\pi, +\pi)$. Functions (297) are readily seen to form an orthogonal and normalized system. For, if $p \neq q$:

$$\int_{-\pi}^{\pi} \overline{\varphi_p(x)}\, \varphi_q(x)\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(q-p)x}\, dx = \frac{1}{2\pi i(q-p)} \left[e^{i(q-p)x}\right]_{x=-\pi}^{x=\pi} = 0,$$

and if $p = q$:

$$\int_{-\pi}^{\pi} |\varphi_p(x)|^2\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx = 1.$$

We can show by using the results obtained earlier for Fourier expansions that system (297) is likewise complete in the interval in question.

51. Linear functional operators. It is possible to establish for functional space a concept corresponding to that of linear transformation for the space H. This leads us to *linear functional operators*. Suppose we have a definite rule by which there corresponds to any function $f(x)$ (with definite properties) another function $F(x)$:

$$F(x) = L[f(x)], \tag{298}$$

L being the symbolic notation for the correspondence rule. Here we have as it were a generalized concept of function. The role of argument is played, not by a variable number, but by a function $f(x)$ which can be chosen arbitrarily from a certain class, whilst again, the value of the function, insteadof being anumber, is a new function $F(x)$. Such a generalized functional relationship is usually described as a *functional operator* or *functional*. The idea of functional operator is present in a latent form in a number of problems of mathematical physics. We may take, for instance, the problem of the vibrations of a string fixed

at the ends. The graph of the string at a given instant t is defined by the two graphs of the initial conditions, i.e. by the graph of the initial displacement and by the graph of initial velocity, so that we are evidently concerned here with a functional operation. The same type of situation is obtained in many other problems of mathematical physics. Sometimes the role of argument is played, not by the graph of an initial distribution, but say by the contour of the domain to which the problem relates.

The operator L is said to be linear if we have

$$L[f_1(x)+f_2(x)] = L[f_1(x)] + L[f_2(x)] \tag{299}$$

and

$$L[cf(x)] = cL[f(x)],$$

where c is a constant.

The condition for the transformation to be bounded is of the form

$$\|L[f(x)]\| \leqslant M\|f(x)\|, \tag{299$_1$}$$

where M is a positive constant and $f(x)$ is any function of the functional space.

We shall not concern ourselves here with the general theory of linear operators but confine the discussion to a few special examples that explain the main essence of the concept; its connection with linear transformations of space H will also be mentioned, since we have established the correspondence between functional space and space H.

In certain cases a linear functional operator may be written as

$$F(x) = \int_a^b K(x, t) f(t)\, dt, \tag{300}$$

where $K(x, y)$ is a given function of two variables which is usually known as the kernel (or nucleus) of the operator. In the present case the kernel is the complete analogue of the array of a_{ik} of a linear transformation of the space H. Instead of subscripts i and k, we have here the two variables x and y which take a continuous sequence of values, and expression (300) is entirely analogous to (266). We shall investigate operators of type (300) in detail when considering integral equations.

Unitary and Hermitian operators may readily be defined for the present case. The linear functional operator L is said to be *unitary* if we have, for any two functions $f(x)$ and $\varphi(x)$ of a given class:

$$(Lf(x), L\varphi(x)) = (f(x), \varphi(x)). \tag{301}$$

The Hermitian operator L_1 is defined by

$$\big(L_1 f(x), \varphi(x)\big) = \big(f(x), L_1 \varphi(x)\big). \tag{302}$$

Let L_1 have the form (300), i.e.

$$L_1 f(x) = \int_a^b K(x, t) f(t)\, dt.$$

We form the scalar products appearing in (302):

$$\big(f(x), L_1 \varphi(x)\big) = \int_a^b \int_a^b \overline{K(x, t)} f(x) \overline{\varphi(t)}\, dt\, dx,$$

$$\big(L_1 f(x), \varphi(x)\big) = \int_a^b \int_a^b K(x, t) f(t) \overline{\varphi(x)}\, dt\, dx.$$

On changing the notation for the variables of integration in the last integral, we can write (302) as follows:

$$\int_a^b \int_a^b [\overline{K(x, t)} - K(t, x)] f(x) \overline{\varphi(t)}\, dt\, dx = 0. \tag{303}$$

If the kernel of the operator satisfies the relationship

$$K(x, t) = \overline{K(t, x)}, \tag{304}$$

condition (303) is satisfied for any choice of functions $f(x)$ and $\varphi(x)$, and the operator L_1 is Hermitian in this case. On taking into account the arbitrariness of the above-mentioned $f(x)$ and $\varphi(x)$, we can say that equation (304) is not only sufficient but also necessary for condition (303) to be fulfilled, i.e. for the operator L_1 to be Hermitian. If the kernel $K(x, t)$ is a real function, condition (304) can be written as

$$K(x, t) = K(t, x), \tag{305}$$

i.e. the kernel must in this case be a symmetric function of its arguments.

We consider a few further examples of linear operators. We take as our first example the operator consisting of differentiation followed by multiplication by $1/i$:

$$Lf(x) = \frac{1}{i} \frac{df(x)}{dx} = \frac{1}{i} f'(x), \tag{306}$$

where $(-\pi, +\pi)$ is taken as the basic interval. We form the scalar product for operator (306):

$$\left(f(x), L\varphi(x)\right) = \left(f(x), \frac{1}{i}\varphi'(x)\right) = \frac{-1}{i}\int_{-\pi}^{\pi} f(x)\overline{\varphi'(x)}\,\mathrm{d}x.$$

On assuming periodic functions of period 2π and integrating by parts, we have

$$\left(f(x), \frac{1}{i}\varphi'(x)\right) = \frac{-1}{i} f(x)\overline{\varphi(x)}\Big|_{x=-\pi}^{x=\pi} + \frac{1}{i}\int_{-\pi}^{\pi} f'(x)\overline{\varphi(x)}\,\mathrm{d}x,$$

whence the equation follows at once:

$$\left(f(x), \frac{1}{i}\varphi'(x)\right) = \left(\frac{1}{i}f'(x), \varphi(x)\right), \tag{307}$$

i.e. (306) is an Hermitian operator with respect to the class of differentiable periodic functions.

We select the system of functions (297) as the coordinate system in functional space. The function $f(x)$ is now characterized by its Fourier coefficients a_k, which are given by

$$a_k = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} \mathrm{e}^{-ikx} f(x)\,\mathrm{d}x. \tag{308}$$

We shall have different Fourier coefficients a'_k for the function $(1/i)f'(x)$. We can readily establish a linear transformation giving the a'_k in terms of the a_k. This linear transformation will express the functional operator (306) in the form of an infinite matrix; at the same time, it must be borne in mind that this expression for operator (306) will be referred to a definite choice of coordinate axes in the functional space, namely the coordinate functions given by (297). We have:

$$a'_k = \frac{1}{\sqrt{2\pi}i}\int_{-\pi}^{\pi} \mathrm{e}^{-ikx} f'(x)\,\mathrm{d}x,$$

whence, on integrating by parts and assuming $f(x)$ periodic, we find:

$$a'_k = \frac{k}{\sqrt{2\pi}}\int_{-\pi}^{\pi} \mathrm{e}^{-ikx} f(x)\,\mathrm{d}x,$$

i.e.

$$a'_k = k a_k \qquad (k = 0, \pm 1, \pm 2, \ldots). \tag{309}$$

This equation in fact expresses the linear transformation concerned. Its matrix has the form

$$\left\|\begin{array}{cccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \ldots -2, & 0, & 0, & 0, & 0, & \ldots \\ \ldots \quad 0, & -1, & 0, & 0, & 0, & \ldots \\ \ldots \quad 0, & 0, & 0, & 0, & 0, & \ldots \\ \ldots \quad 0, & 0, & 0, & 1, & 0, & \ldots \\ \ldots \quad 0, & 0, & 0, & 0, & 2, & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array}\right\|. \tag{310}$$

i.e. the matrix is seen to be diagonal. The fact that the rows and columns of (310) are numbered from $-\infty$ to $+\infty$ instead of from 1 to ∞ is a new point of no real significance. Functions (297) are numbered in the same manner. We remark that these functions satisfy the obvious relationship

$$\frac{1}{i}\varphi_k'(x) = k\varphi_k(x),$$

i.e. on writing L for operator (306):

$$L\varphi_k(x) = k\varphi_k(x). \tag{311}$$

By analogy with [37], we can call the $\varphi_k(x)$ the *eigenfunctions of the operator* L and the k the *corresponding eigenvalues.* The diagonal form of matrix (310) is directly bound up with the fact that the $\varphi_k(x)$ are the eigenfunctions of operator (306).

We take as a second example the operator consisting of multiplication by the independent variable:

$$L_1[f(x)] = xf(x). \tag{312}$$

We find the linear transformation expressing this operator in the space H, when we take (297) as coordinate functions in the functional space. Let a_k be the Fourier coefficients of $f(x)$ as above, and a_k' the Fourier coefficients of $xf(x)$, i.e.

$$a_m' = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} e^{-imx} xf(x)\,dx \qquad (m = 0, \pm 1, \ldots). \tag{313}$$

We want to find the linear transformation giving a_m' in terms of a_m.

To evaluate integral (313), we find the Fourier coefficients of $(1/\sqrt{2\pi})\, e^{imx}x$:

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(m-k)x}\, x\,dx.$$

Integration by parts gives us, for $m - k \neq 0$:

$$c_k = \frac{1}{i\,(m-k)}\, e^{i\,(m-k)\pi} = \frac{(-1)^{m-k}}{i\,(m-k)}\,.$$

We now find the coefficient c_m:

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} x \, dx = 0\,.$$

We re-write (313) as

$$c'_m = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \overline{(e^{imx}\, x)}\, f(x)\, dx,$$

and use the generalized closure equation (286), the Fourier coefficients of $f(x)$ being given by a_k and the Fourier coefficients of $(1/\sqrt{2\pi})\, e^{imx}\, x$ by the above expressions. We obtain for a'_m:

$$a'_m = i \sum_{k=-\infty}^{+\infty}{}' \frac{(-1)^{m-k}}{m-k}\, a_k\,; \tag{314}$$

the prime on the summation sign indicates that terms corresponding to $k = m$ must be excluded. Expression (314) in fact gives the linear transformation of space H corresponding to operator (312), if (297) are taken as the coordinate functions in the functional space.

In general, suppose we take as coordinate functions some complete system of orthogonal and normalized functions

$$\varphi_1(x), \varphi_2(x), \ldots$$

If H is a linear Hermitian operator, where

$$H\varphi_i(x) \sim \sum_{k=1}^{\infty} a_{ki}\varphi_k(x),$$

we have $a_{ik} = \overline{a_{ki}}$. Let

$$\psi(x) \sim \sum_{k=1}^{\infty} c_k \varphi_k(x)$$

give the Fourier series of a function $\psi(x)$. For the function $H\psi(x)$, we have the new Fourier series

$$H\psi(x) \sim \sum_{k=1}^{\infty} c'_k \varphi_k(x)\,,$$

where it can be shown that

$$c'_k = \sum_{j=1}^{\infty} a_{kj} c_j \qquad (k = 1,\ 2,\ \ldots).$$

This linear transformation in fact expresses the operator H if the $\varphi_k(x)$ are taken as coordinate functions.

We return to the differentiation operator (306). Even if we only consider continuous functions, this operator cannot be applied to all of them, since continuous functions exist that lack a derivative for every value of x. The linear transformation (309) of space H corresponds to operator (306). If the series $\sum_{-\infty}^{+\infty} |a_k|^2$ is convergent, the series

$$\sum_{-\infty}^{+\infty} |a_k'|^2 = \sum_{-\infty}^{+\infty} k^2 |a_k|^2$$

may conceivably be divergent. This shows that transformation (309) is not applicable to the whole of the space H, which is in accordance with what we said above.

CHAPTER III

THE BASIC THEORY OF GROUPS AND LINEAR REPRESENTATIONS OF GROUPS

52. Groups of linear transformations. We consider the set of all unitary transformations in n-dimensional space. All these transformations have a non-zero determinant, so that for any unitary transformation $U\mathbf{x}$ which is completely characterized by its matrix U there is a fully defined inverse transformation $U^{-1}\mathbf{x}$ which is also unitary [28]. Furthermore, if $U_1\mathbf{x}$ and $U_2\mathbf{x}$ are two unitary transformations, their product $U_2 U_1 \mathbf{x}$ is also unitary. All these properties of the set of all unitary transformations may be briefly expressed by saying that *the set of unitary transformations forms a group.*

A set of linear transformations with non-zero determinants in general forms a group if the following two conditions are fulfilled: firstly, *the inverse of any transformation belonging to the set also belongs to the set,* and secondly, *the product in any order of two transformations belonging to the set also belongs to the set, the transformations multiplied being possibly identical.*

Bearing in mind that the product of any transformation with its inverse is the identity transformation, we can say that *a group must contain the identity transformation,* i.e. *the unit matrix.*

Since a linear transformation is always fully defined by its matrix, it is immaterial whether, in the above or in what follows, we speak of *groups of linear transformations or groups of matrices.*

Further examples may be given of groups of linear transformations. The set of all real orthogonal transformations may easily be seen to form a group. We know that real orthogonal transformations have determinants equal to (± 1). If we take the set of real orthogonal transformations with $(+1)$ determinants, we also get a group. The set of real orthogonal transformations with (-1) determinants do not

form a group, however, since the product of two matrices with (-1) determinants yields a matrix with a $(+1)$ determinant.

In particular, if we take the group of real orthogonal transformations in three variables, this consists of pure rotations of space about the origin, and of the transformations resulting from such a rotation together with a symmetry transformation with respect to the origin. Whereas if we take the group of linear orthogonal transformations in three variables with $(+1)$ determinants, we get the group of rotations of space about the origin.

All the groups mentioned have contained an infinite set of transformations; in particular, the group of rotations of three-dimensional space about the origin depends on three arbitrary real parameters, i.e. the Euler's angles that we discussed above.

We take as an example the rotation of space about the z axis by an angle φ, the expressions for which are

$$\left.\begin{aligned} x' &= x\cos\varphi - y\sin\varphi \\ y' &= x\sin\varphi + y\cos\varphi. \end{aligned}\right\} \tag{1}$$

If the real parameter φ takes all values in the interval $(0, 2\pi)$, we obviously get a group containing an infinite set of transformations and depending on a single real parameter. We bring in the following notation for the matrix of the transformation:

$$Z_\varphi = \begin{Vmatrix} \cos\varphi, & -\sin\varphi \\ \sin\varphi, & \cos\varphi \end{Vmatrix}. \tag{2}$$

It is immediately clear that the product of two rotations by the angles φ_1 and φ_2 yields a rotation by the angle $(\varphi_1 + \varphi_2)$:

$$Z_{\varphi_2} Z_{\varphi_1} = Z_{\varphi_2+\varphi_1} \tag{3}$$

and similarly,

$$Z_{\varphi_1} Z_{\varphi_2} = Z_{\varphi_1+\varphi_2}.$$

This shows us that here all the transformations, or as we say, all the elements of the group commute in pairs. Such a group is termed *Abelian*. In the last example, moreover, multiplication of two elements amounts simply to addition of the two corresponding parameters.

We can somewhat extend the last example by taking optical reflection in the y axis as well as rotation of the xy plane about the origin. It is clearly immaterial in what order these operations are carried out, i.e. we can first rotate about the origin then reflect symmetrically in the y axis, or vice versa. Changing the order affects the result, but the

total set of transformations is the same in both cases. The set consists of real orthogonal transformations in two variables. The matrix has the general form

$$\{\varphi, d\} = \begin{Bmatrix} d\cos\varphi, & -d\sin\varphi \\ \sin\varphi, & \cos\varphi \end{Bmatrix}, \tag{4}$$

where φ is the previous parameter and d is a number equal to ± 1. With $d = 1$ we get rotation of the xy plane about the origin, whilst with $d = -1$ we get rotation followed by the reflection. The following rule is readily derived for multiplication of matrices (4):

$$\{\varphi_2 d_2\}\{\varphi_1, d_1\} = \{\varphi_1 + d_1\varphi_2, d_1 d_2\}. \tag{5}$$

The product can now depend on the order of the factors, i.e. the present group is not Abelian. Similarly, the group of real orthogonal transformations in three-dimensional space is clearly not Abelian, and the same can be said even of the group of rotations of three-dimensional space.

The examples so far have been of groups containing an infinite set of transformations (elements), the corresponding matrices having contained arbitrary real parameters. We now mention some examples of groups containing a finite number of elements. Let m be a given positive integer. We consider the set of rotations of the xy plane about the origin by the angles

$$0, \frac{2\pi}{m}, \frac{4\pi}{m}, \ldots, \frac{2(m-1)\pi}{m}.$$

so that we have altogether m transformations, with matrices

$$Z_{\frac{2k\pi}{m}} = \begin{Vmatrix} \cos\frac{2k\pi}{m}, & -\sin\frac{2k\pi}{m} \\ \sin\frac{2k\pi}{m}, & \cos\frac{2k\pi}{m} \end{Vmatrix} \quad (k = 0, 1, \ldots, m-1).$$

These transformations clearly form a group, the elements of which are positive integral powers of a single transformation, i.e.

$$Z_{\frac{2k\pi}{m}} = \left(Z_{\frac{2\pi}{m}}\right)^k \quad (k = 0, 1, \ldots, m-1). \tag{6}$$

A finite group made up of powers of a transformation is usually described as *cyclical*.

If we take an angle φ_0, not a multiple of π, the transformations (matrices)

$$Z_{\varphi_0}^k = Z_{k\varphi_0} \quad (k = 0 \pm 1, \pm 2, \ldots) \tag{7}$$

clearly also form a group. But we now have a group with an infinite set of elements, since there is no integral power for which $Z^k_{\varphi_0}$ coincides with $Z^0_{\varphi_0} = I$. Group (7) is infinite, yet its matrices do not contain a continuously varying parameter. We say in this case that *the elements in the group are enumerable*, i.e. we can provide every element with an integral subscript, in such a way that different subscripts correspond to different elements, and every integer is the subscript of an element. We cannot do this in the case of groups containing continuously varying parameters.

53. Groups of regular polyhedra. Finite groups may also be formed by rotation of three-dimensional space about the origin. We know that these rotations are expressed, in a given coordinate system, by linear transformations of the coordinates. It must be pointed out that when we speak of a rotation of space about the origin, we understand simply the final effect of passage from the initial to the transformed position. How the passage is achieved is completely immaterial for our discussion. In fact, any linear transformation defines the coordinates of transformed points but naturally says nothing about the actual mechanism of the transformation, so that this latter plays no part at all in our arguments.

We take a sphere with centre at the origin and unit radius. We inscribe a regular polyhedron in the sphere, say the octahedron of Fig. 2. The surface of this polyhedron is known to consist of eight equilateral triangles. We consider the set of rotations of three-dimensional space about the origin such that the octahedron is transformed into itself. This set may easily be seen to form a group which contains a finite number of elements. Let us find the number of elements. We take any axis joining two opposite vertices of the octahedron. The octahedron is transformed into itself if we rotate the space about this axis by angles of 0, $\pi/2$, π, $3\pi/2$. Rotation by the angle 0 is evidently the identity transformation,

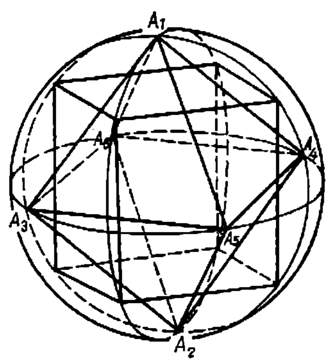

FIG. 2

i.e. corresponds to the unit matrix. We shall write our four rotations about this axis as

$$S_0 = I, S_1, S_2, S_3. \tag{8}$$

Let A be a vertex of the octahedron on our axis. We introduce in the five linear transformations

$$T_1, T_2, T_3, T_4, T_5,$$

transforming the octahedron into itself and such that A coincides with one of the remaining five vertices. Along with the four rotations (8) we compose a further 20 rotations of space about the origin as follows:

$$T_k S_0, \ T_k S_1, \ T_k S_2, \ T_k S_3 \qquad (k = 1, \ 2, \ 3, \ 4, \ 5). \tag{9}$$

The 24 rotations (8) and (9) are easily seen to be distinct. This is completely obvious geometrically, whilst it can also be shown as follows: let

$$T_p S_q = T_{p_1} S_{q_1}. \tag{10}$$

The transformations S_i correspond to rotations about an axis passing through the vertex A, so that they do not alter the position of A. Transformations T_p and T_{p_1}, for different subscripts p and p_1, shift A to different vertices, and it therefore follows from (10) that p and p_1 are the same; but now it follows from the same equation, after multiplying on the left by $T_p^{-1} = T_{p_1}^{-1}$, that q and q_1 are the same, i.e. (10) is only valid when the left and right-hand sides consist of the same factors. We thus have 24 distinct rotations from (8) and (9), for which the octahedron is displaced into itself. We now show that these are all the rotations having this property. Let V be a rotation transforming the octahedron into itself. Suppose that, with this, A is displaced to another vertex A_j, and let T_j be the transformation T_k which also shifts A to A_j. We compose the transformation $T_j^{-1} V$. With this, the octahedron transforms into itself and A remains in position. The opposite vertex consequently also remains in position, so that the transformation we have composed is one of the rotations S_i about the axis through A, i.e. $T_j^{-1} V = S_i$, whence $V = T_j S_i$. In other words, any transformation shifting the octahedron into itself must be included in the 24 rotations composed above. Or finally, *the group of rotations transforming an octahedron into itself consists of 24 elements.*

We can evidently inscribe a cube in the unit sphere such that the radii passing through the centres of the faces of the octahedron end in the vertices of the cube. It follows directly from this that the group of rotations is the same for a cube as for as octahedron. Suppose we take a new position of the octahedron, obtained from the original position with the aid of a rotation having the matrix U. If V is a rota-

tion that displaces the original octahedron into itself, UVU^{-1} clearly yields a rotation that displaces the new octahedron into itself, and conversely. Thus if the group of rotations of the original octahedron consists of matrices V_k $(k = 1, 2, \ldots, 24)$, the rotation group for the new octahedron simply consists of the similar matrices $UV_k U^{-1}$. In other words, we obtain a *similar group*. In general, *if a set of matrices V_k forms a group, the set of similar matrices $UV_k U^{-1}$, with any fixed U, also forms a group.* We leave to the reader the proof, that easily follows directly from the definition of group. The second group is usually described as *similar* to the first.

We now consider the tetrahedron, having four vertices and a surface consisting of four equilateral triangles. We take any axis joining a vertex A to the centre of the opposite face. The tetrahedron is transformed into itself if we rotate the space in a given direction about the axis by angles of 0, $2\pi/3$, $4\pi/3$. Let these rotations be S_0, S_1, S_2. We also introduce the three linear transformations T_1, T_2, T_3 by which the tetrahedron is displaced into itself, the vertex A being brought to coincide with one of the three remaining vertices. In addition to S_0, S_1, S_2, we compose the nine rotations $T_k S_0$, $T_k S_1$, $T_k S_2$ $(k = 1, 2, 3)$ which gives us altogether 12 rotations that are distinct and that represent all the rotations transforming the tetrahedron into itself.

We now take the icosahedron whose surface consists of twenty equilateral triangles and whose vertices are twelve in number. As above, we take any axis joining a vertex A to the opposite vertex. The icosahedron is displaced into itself by rotating the space by angles of $2k\pi/5$ $(k = 0, 1, 2, 3, 4)$. Let these rotations be S_k. We have further eleven rotations T_l $(l = 1, 2, \ldots, 11)$ for which the vertex A becomes one of the remaining vertices and the icosahedron is displaced into itself. The total group of rotations transforming the icosahedron into itself consists of the five rotations S_k and 55 rotations $T_l S_k$. Thus the group contains altogether 60 rotations. The same group is obtained for a dodecahedron, with twenty vertices and twelve regular pentagonal faces. This can be seen by arranging the dodecahedron with respect to the icosahedron in a similar manner to that used above for arranging a cube with respect to an octahedron.

We consider one further group of rotations of three-dimensional space. Suppose we have a regular n-sided polygon in the xy plane, with its centre at the origin. We take an axis joining a vertex A to the opposite vertex (if n is even), or to the middle of the opposite

side (if n is odd). The polygon is displaced into itself by rotation about the axis by angles of 0 and π. The first rotation is the identity transformation I, whilst we write S for the second.

In addition, the rotations T_k about the z axis by angles of $2k\pi/n$ ($k = 1, 2, \ldots, n-1$) displace the vertex A to another vertex and transform the polygon into itself. We have the identity transformation $T_0 = I$ with $k = 0$. The total group of transformations displacing the polygon into itself contains the following $2n$ elements: T_k and $T_k S$ ($k = 0, 1, 2, \ldots, n-1$).

The above n-sided polygon, whose surface is taken twice (top and bottom), is usually termed a dihedron, the corresponding group being the dihedral group.

54. Lorentz transformations. All the above examples of groups of linear transformations have consisted of unitary transformations or of rotations of three-dimensional space (a particular case of unitary transformation). We now investigate a new group of linear transformations where the elements are not unitary. This group has an important role in relativity, electrodynamics and relativistic quantum mechanics.

We take four variables x_1, x_2, x_3, x_4, the first three being the spatial coordinates of a point and the last being time. The fundamental requirement of the special theory of relativity for the invariance of a certain definite velocity c (the speed of light) in the case of relative motion leads to the following problem: for what linear transformations of the above four variables is the expression

$$x_1^2 + x_2^2 + x_3^2 - c^2x_4^2$$

invariant? To be more explicit, we want to find the linear transformations giving the new variables x_k' in terms of the original x_k such that we have the identity

$$x_1'^2 + x_2'^2 + x_3'^2 - c^2x_4'^2 = x_1^2 + x_2^2 + x_3^2 - c^2x_4^2.$$

We first take the case when the coordinates x_2 and x_3 remain unchanged, so that x_1 and x_4 are the only variables in the linear transformation. We thus have to find the transformations

$$x_1' = a_{11}x_1 + a_{14}x_4, \quad x_4' = a_{14}x_1 + a_{44}x_4, \tag{11}$$

such that

$$x_1'^2 - c^2x_4'^2 = x_1^2 - c^2x_4^2. \tag{12}$$

We replace x_4 by a new pure imaginary variable given by

$$y_1 = icx_4.$$

The required linear transformations must have the form

$$x_1' = \alpha_{11}x_1 + \alpha_{12}y_1, \quad y_1' = \alpha_{21}x_1 + \alpha_{22}y_1, \tag{13}$$

where

$$\alpha_{11} = a_{11}; \quad \alpha_{12} = \frac{a_{14}}{ic}; \quad \alpha_{21} = ica_{41}; \quad \alpha_{22} = a_{44},$$

whilst condition (12) may be re-written as

$$x_1'^2 + y_1'^2 = x_1^2 + y_1^2. \tag{14}$$

The coefficients α_{11} and α_{22} must be real, and α_{12} and α_{21} pure imaginary. We therefore write $\alpha_{12} = i\beta_{12}$ and $\alpha_{21} = i\beta_{21}$. Condition (14) is evidently equivalent to requiring orthogonality of transformations (13), and we can write that the sum of the squares of the elements of each row and column must be equal to unity. It is easily verified that this gives us $\beta_{12} = \beta_{21} = \alpha_{11} - 1 = \alpha_{22} - 1$ and $\alpha_{11}^2 = \alpha_{22}^2$. Let $\alpha_{22} = a$ and $\beta_{12} = a\beta$. We shall take α_{11} and α_{22} positive, which corresponds to invariance of the direction of measuring x_1 and x_4. Thus instead of (13), we have by the above relationships:

$$x_1' = ax_1 + ia\beta y_1, \quad y_1' = \alpha_{21}x_1 + ay_1.$$

The condition for orthogonality of rows,

$$a\alpha_{21} + ia^2\beta = 0,$$

gives us $\alpha_{21} = -ia$, i.e. β_{12} and β_{21} must have opposite signs. Finally the condition

$$\alpha_{11}^2 + \alpha_{12}^2 = 1$$

gives us

$$a^2 - a^2\beta^2 = 1, \qquad a = \frac{1}{\sqrt{1-\beta^2}} \qquad (\beta^2 < 1),$$

and we arrive at the following expressions:

$$x_1' = \frac{x_1 + i\beta y_1}{\sqrt{1-\beta^2}}; \quad y_1' = \frac{-i\beta x_1 + y_1}{\sqrt{1-\beta^2}},$$

or, on again returning from $y_1 = icx_4$ to the original x_4:

$$x_1' = \frac{x_1 - \beta c x_4}{\sqrt{1-\beta^2}}; \quad x_4' = \frac{-\dfrac{\beta}{c}x_1 + x_4}{\sqrt{1-\beta^2}}. \tag{15}$$

It follows directly from these equations that the coordinate system corresponding to the primed variables moves with respect to the original coordinate system with a velocity

$$v = \beta c \tag{16}$$

in the direction of the x_1 axis. For if we take x_1' constant, we get

$$\mathrm{d}x_1 - \beta c\,\mathrm{d}x_4 = 0, \quad \text{i.e.} \quad \frac{\mathrm{d}x_1}{\mathrm{d}x_4} = \beta c.$$

On replacing β by the velocity v in accordance with (16), x_1 by x, and x_4 by t, we get the usual form of the Lorentz transformation in two variables:

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}; \quad t' = \frac{-\frac{v}{c^2}x + t}{\sqrt{1 - \frac{v_2}{c^2}}}. \tag{17}$$

In the limit as $c \to \infty$, we get the ordinary expressions for relative motion in classical mechanics:

$$x' = x - vt; \quad t' = t.$$

The Lorentz transformations (17), depending on the single real parameter v, may easily be seen to form a group. On solving equations (17) with respect to x and t, we obtain the inverse transformation to (17). This is the Lorentz transformation obtained by replacing v by $(-v)$ in (17). For we have by solving (17):

$$\left(1 - \frac{v^2}{c^2}\right)x = \sqrt{1 - \frac{v^2}{c^2}}\,(x' - vt'); \ \left(1 - \frac{v^2}{c^2}\right)t = \sqrt{1 - \frac{v^2}{c^2}}\left(\frac{v}{c^2}x' + t'\right),$$

whence it follows that

$$x = \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}}; \ t = \frac{\frac{v}{c^2}x' + t'}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

We now consider the Lorentz transformations L_1 and L_2 corresponding to the parametric values $v = v_1$ and $v = v_2$. We form their product $L_2 L_1$ and show that this is also a Lorentz transformation. We have to form the product of two matrices:

$$\left\| \begin{matrix} \frac{1}{\sqrt{1-\beta_2^2}}, & -\frac{\beta_2 c}{\sqrt{1-\beta_2^2}} \\ -\frac{\frac{\beta_2}{c}}{\sqrt{1-\beta_2^2}}, & \frac{1}{\sqrt{1-\beta_2^2}} \end{matrix} \right\| \quad \left\| \begin{matrix} \frac{1}{\sqrt{1-\beta_1^2}}, & -\frac{\beta_1 c}{\sqrt{1-\beta_1^2}} \\ -\frac{\frac{\beta_1}{c}}{\sqrt{1-\beta_1^2}}, & \frac{1}{\sqrt{1-\beta_1^2}} \end{matrix} \right\|,$$

where

$$\beta_1 = \frac{v_1}{c}; \quad \beta_2 = \frac{v_2}{c}.$$

The usual multiplication rules give us the following matrix product

$$\frac{1+\beta_1\beta_2}{\sqrt{1-\beta_2^2}\sqrt{1-\beta_1^2}} \begin{Vmatrix} 1, & -\dfrac{\beta_1 c + \beta_2 c}{1+\beta_1\beta_2} \\ -\dfrac{\dfrac{\beta_1}{c}+\dfrac{\beta_2}{c}}{1+\beta_1\beta_2}, & 1 \end{Vmatrix}. \tag{18}$$

We introduce the new quantity

$$v_3 = \frac{v_1+v_2}{1+\dfrac{v_1 v_2}{c^2}}. \tag{19}$$

We can easily verify the identity

$$\frac{1+\dfrac{v_1 v_2}{c^2}}{\sqrt{1-\dfrac{v_2^2}{c^2}}\sqrt{1-\dfrac{v_1^2}{c^2}}} = \frac{1}{\sqrt{1-\dfrac{v_3^2}{c^2}}},$$

as a result of which matrix (18) can be written as follows:

$$\begin{Vmatrix} \dfrac{1}{\sqrt{1-\beta_3^2}}, & -\dfrac{\beta_3 c}{\sqrt{1-\beta_3^2}} \\ -\dfrac{\dfrac{\beta_3}{c}}{\sqrt{1-\beta_3^2}}, & \dfrac{1}{\sqrt{1-\beta_3^2}} \end{Vmatrix} \qquad \left(\beta_3 = \frac{v_3}{c}\right),$$

i.e. it in fact corresponds to the Lorentz transformation for $v = v_3$. Expression (19) thus gives the *rule for adding velocities in the special theory of relativity*. If we set $v_1 = c$ in (19), it may readily be seen that we also get $v_3 = c$ for the resultant velocity, i.e. *the velocity c is in fact unchanged when two motions are superimposed.*

When deriving (15), we fixed the signs of the coefficients of linear transformation (11) in a definite manner, i.e. we assumed a_{11} and a_{44} positive. An alternative requirement is positiveness of the coefficient a_{44} and of the determinant

$$a_{11}a_{41} - a_{14}a_{41}. \tag{20}$$

Positiveness of a_{11} is easily seen to follow as a consequence of this, and vice versa. For the determinant of transformation (17) is equal to $(+1)$, i.e. with $a_{11} > 0$, (20) is also positive. If we were to

take $a_{11} = -a$ and $a_{44} = a$, where $a > 0$, we should have a transformation with a (-1) determinant. The condition that a_{11} is positive is equivalent to the fact that we have $x'_4 \to \infty$ with fixed x_1 and $x_4 \to \infty$. This can be said to correspond to invariance in the direction for measuring time. Thus the formulae do not give all the linear transformations satisfying condition (12), but only those for which determinant (20) is positive and which do not vary the direction for measuring time.

We now return to the *general Lorentz transformation* for four variables x_k $(k = 1, 2, 3, 4)$, where we must satisfy the condition

$$x_1'^2 + x_2'^2 + x_3'^2 - c^2 x_4'^2 = x_1^2 + x_2^2 + x_3^2 - c^2 x_4^2. \tag{21}$$

We shall take x_k $(k = 1, 2, 3)$ and x'_k $(k = 1, 2, 3)$ as Cartesian coordinates in two different three-dimensional spaces R and R'. We show that, by suitable choice of coordinate axes in the two spaces, the general Lorentz transformation can be reduced to the particular case discussed above. Let T denote the general and S the particular Lorentz transformation. Our assertion is equivalent to the fact that we can write T as

$$T = VSU, \tag{22}$$

where U and V are real orthogonal transformations corresponding to the above-mentioned coordinate transformations in spaces R and R'.

We bring in four new variables as above:

$$y_1 = x_1;\ y_2 = x_2;\ y_3 = x_3;\ y_4 = icx_4,$$

and similarly

$$y'_1 = x'_1;\ y'_2 = x'_2;\ y'_3 = x'_3;\ y'_4 = icx'_4.$$

We obtain for the new variables, instead of (21), the ordinary condition for an orthogonal transformation:

$$y_1'^2 + y_2'^2 + y_3'^2 + y_4'^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2. \tag{23}$$

The required linear transformation will be of the form

$$y'_k = a_{k1} y_1 + a_{k2} y_2 + a_{k3} y_3 + a_{k4} y_4 \quad (k = 1, 2, 3, 4). \tag{24}$$

Observing that y_4 and y'_4 must be pure imaginary, we can say that coefficients a_{k1}, a_{k2}, a_{k3}, with $k = 1, 2, 3$, and also a_{44} must be real, whilst a_{41}, a_{42}, a_{43}, and a_{k4} with $k = 1, 2, 3$ must be pure imaginary. A change of coordinate axes in the space R' is equivalent to a real orthogonal transformation on the variables y'_1, y'_2, y'_3. We consider the coefficients

$$a_{14} = i\beta_{14};\ a_{24} = i\beta_{24};\ a_{30} = i\beta_{34}.$$

The real numbers $\beta_{14}, \beta_{24}, \beta_{34}$ define a certain vector; if we take the direction of this as the new first axis in the space R', the coefficients a_{24} and a_{34} vanish as a result of the corresponding orthogonal transformation. To see this, we only need to notice that by (24), an orthogonal transformation on the variables y_1', y_2', y_3' amounts to the same transformation on $\beta_{14}, \beta_{24}, \beta_{34}$. We shall, therefore, suppose that this coordinate transformation in the space R' has already been carried out, so that we have $a_{24} = a_{34} = 0$. Condition (23) shows that the coefficients of transformation (24) must satisfy the ordinary conditions for an orthogonal transformation. On recalling that the coefficients mentioned vanish, consideration of the second and third rows gives us the following conditions:

$$a_{k1}^2 + a_{k2}^2 + a_{k3}^2 = 1, \quad (k = 2, 3)$$
$$a_{21} a_{31} + a_{22} a_{32} + a_{23} a_{33} = 0,$$

where all the coefficients present are real. By the conditions written, the vectors with components (a_{21}, a_{22}, a_{23}) and (a_{31}, a_{32}, a_{33}) are mutually orthogonal and of unit length. If we choose these two as fundamental vectors in the space R, directed along the x_2 and x_3 axes, the two sums

$$a_{k1} y_1 + a_{k2} y_2 + a_{k3} y_3 \quad (k = 2, 3),$$

expressing the scalar products of the two vectors with the variable vector (y_1, y_2, y_3), reduce simply to the forms y_2 and y_3, i.e. with this choice of coordinate axes we have:

$$a_{22} = a_{33} = 1; \quad a_{21} = a_{23} = a_{31} a_{32} = 0.$$

With the axes chosen in the two spaces, the matrix of transformation (24) takes the form

$$\begin{Vmatrix} a_{11}, a_{12}, a_{13}, a_{14} \\ 0, \ 1, \ 0, \ 0 \\ 0, \ 0, \ 1, \ 0 \\ a_{41}, a_{42}, a_{43}, a_{44} \end{Vmatrix}. \tag{25}$$

This matrix has been obtained as a result of multiplying the original matrix by two orthogonal transformations of only the first three variables, though they can evidently be looked on as transformations of the four variables, the fourth variable being kept constant. Since the product of two orthogonal transformations must also be orthogonal, we can say that the elements of (25) also satisfy the

orthogonality condition. On writing down this condition for the first row with respect to the second and third, we get

$$a_{12} = a_{13} = 0,$$

and similarly, for the fourth row with respect to the second and third:

$$a_{43} = a_{42} = 0.$$

We thus arrive at the following matrix:

$$\begin{Vmatrix} a_{11}, 0, 0, a_{14} \\ 0, \quad 1, 0, 0 \\ 0, \quad 0, 1, 0 \\ a_{41}, 0, 0, a_{44} \end{Vmatrix},$$

i.e. we have in this case the linear transformation

$$\begin{aligned} y_1' &= a_{11} y_1 + a_{14} y_4, \\ y_4' &= a_{41} y_1 + a_{44} y_4 \end{aligned}$$

which has to satisfy the condition

$$y_1'^2 + y_4'^2 = y_1^2 + y_4^2.$$

We dealt with this transformation above; it led us to the special Lorentz transformation (15), and (22) can thus be taken as proved. We just notice that the sign rule is the same in defining transformation S as previously, if the general Lorentz transformation T is required not to change the direction for measuring time and also to have a determinant greater than zero. We can always regard the orthogonal transformations U and V as rotations of three-dimensional space, so that their determinant will be greater than zero; at the same time, they in no way affect the fourth variable. We can thus conclude that transformation S must also have a determinant greater than zero, whilst it must not affect the measurement of time, i.e. given our assumption regarding the general transformation T, we arrive at precisely the conditions for the special transformation under which our formulae were derived. The general transformation satisfying the two conditions postulated above is usually termed a *positive Lorentz transformation*. It follows from the above discussion that the matrices corresponding to these are given by (22), where S is the particular Lorentz transformation of type (15) and U and V are the matrices of rotations of three-dimensional space. Positive Lorentz transformations, like transformations (15), can be shown to form a group.

The above arguments show that the matrix of the most general Lorentz transformation, defined only by condition (21), can be written in the form (22), where U and V are rotations and S is the general Lorentz transformation in two variables. If this is a positive transformation, it follows at once from (15) that $D(S) = 1$, and the determinant of every positive Lorentz transformation is also equal to unity, inasmuch as the determinants of U and V are unity, matrices U, S, and V being considered as of the fourth order. As may easily be seen, the determinant can be equal to (± 1) in the general case of a second order Lorentz transformation, so that the general transformation will likewise have a determinant of (± 1).

55. Permutations. We have so far considered examples of groups whose elements are linear transformations. There is no essential connection between the group concept and linear transformations, and we can construct groups for other types of operation. Our next discussion concerns *permutations*, representing a type of operation that we have already encountered in [2]. We must first mention some basic facts and concepts in regard to permutations.

Suppose we have n objects, which we enumerate as in [2], i.e. we can simply suppose that the objects are the integers $1, 2, \ldots n$. As we know, $n!$ permutations are possible of these numbers. We take one such permutation:

$$p_1, p_2 \cdots, p_n. \tag{26}$$

This set of p_k yields all the numbers from 1 to n, arranged in a definite order in accordance with (26). We compare permutation (26) with the basic permutation $1, 2, \ldots, n$:

$$\begin{pmatrix} 1, & 2, & \ldots, & n \\ p_1, & p_2, & \ldots, & p_n \end{pmatrix} (P). \tag{27}$$

The passage from the basic permutation to (26) is accomplished by replacing 1 by p_1, 2 by p_2, and so on. We denote this operation by the single letter P and refer to it in future as a permutation. We now define the *inverse permutation* P^{-1}. This is the operation such that (26) becomes the basic permutation, i.e. p_1 is replaced by 1, p_2 by 2, and so on. We can explain this by means of an example: suppose $n = 5$ and we have the permutation

$$\begin{pmatrix} 1, 2, 3, 4, 5 \\ 3, 2, 5, 1, 4 \end{pmatrix} (P).$$

The inverse permutation will be

$$\begin{pmatrix}1, 2, 3, 4, 5\\ 4, 2, 1, 5, 3\end{pmatrix}(P^{-1}).$$

It is easily seen that

$$(P^{-1})^{-1} = P. \tag{28}$$

We now introduce the *product of permutations*. Let P_1 and P_2 be any two permutations. Their product $P_2 P_1$ is defined as the result of carrying out first the permutation P_1 and then P_2. For instance, if we have the two permutations

$$\begin{pmatrix}1, 2, 3, 4, 5\\ 5, 1, 4, 3, 2\end{pmatrix}(P_2) \quad \text{and} \quad \begin{pmatrix}1, 2, 3, 4, 5\\ 3, 1, 5, 2, 4\end{pmatrix}(P_1),$$

their product $P_2 P_1$ gives the permutation

$$\begin{pmatrix}1, 2, 3, 4, 5\\ 4, 5, 2, 1, 3\end{pmatrix}(P_2 P_1).$$

Obviously, the inverse permutation P^{-1} is fully defined by the condition

$$P^{-1} = PP^{-1} = I, \tag{29}$$

where I denotes the identity permutation, in which each element is replaced by itself.

We can define the product of any number of permutations by applying them successively. A product evidently satisfies the law of association, e.g.

$$P_3 (P_2 P_1) = (P_3 P_2) P_1. \tag{30}$$

For we can either first form the product of P_1 with P_2, then form the product of this with P_3, or else we can replace the successive application of P_2 and P_3 by the application of the single permutation $(P_3 P_2)$ which is equivalent to P_2 and P_3 applied successively. We finally notice that the identity permutation clearly satisfies

$$IP = PI = P, \tag{31}$$

where P is any permutation. Products of permutations do not in general satisfy the commutative law, i.e. the permutations $P_2 P_1$ and $P_1 P_2$ are in general different. We suggest that this be verified for the above example.

We have thus established the basic ideas of inverse and identity permutations and products exactly as was done previously for linear

transformations (matrices). We can now continue the analogy and establish the further concept of *group*. A set of permutations forms a group if the following two conditions are fulfilled: firstly, if a permutation belongs to our set, its inverse also belongs to the set, and secondly, the product in any order of two permutations belonging to the set also belongs to the set. As in the case of linear transformations, the identity permutation necessarily belongs to the set.

The set of all $n!$ permutations clearly forms a group. We now establish the existence of another group consisting only of part of the above. We observe for this that any permutation can be obtained with the aid of transpositions [2], different numbers of these being possible for a given permutation though the number is always even or always odd. The permutations resulting from an even number of transpositions themselves form a group. The group formed by all the permutations is usually termed *symmetric*, whilst the even permutations, i.e. those resulting from an even number of transpositions, form an *alternating group*.

We now consider permutations of a special type. Let $l_1, l_2, \ldots, l_m$ be any m different integers from the integers up to n. Suppose our permutation consists in replacing l_1 by l_2, l_2 by l_3, $\ldots$, l_{m-1} by l_m, and finally, l_m by l_1. A permutation of this type is called a *cycle* and written $(l_1, l_2, \ldots, l_m)$. Cyclic permutations of the numbers inside the brackets give us the cycles

$$(l_2, l_3, \ldots, l_m, l_1), (l_3, l_4, \ldots, l_m, l_1, l_2), \ldots$$

which evidently yield the same permutation as $(l_1, l_2, \ldots, l_m)$. If $m = 1$, i.e. we have the cycle (l_1), the cycle is clearly equivalent to the identity permutation, and there is no point in considering it. A cycle of two numbers (l_1, l_2) is obviously equivalent to transposition of the elements l_1 and l_2.

If we have two cycles with no common elements, their product is independent of the order of the factors.

Suppose say $n = 5$, and we take the product of the two cycles with no common elements,

$$(1, 3)(2, 4, 5) \text{ and } (2, 4, 5)(1, 3).$$

Both these products clearly yield the same permutation

$$\begin{pmatrix} 1, 2, 3, 4, 5 \\ 3, 4, 1, 5, 2 \end{pmatrix}.$$

We can represent any permutation P as a *product of cycles* having no common elements. To do this, we take the element 1 as the first element in a cycle. We take as the second element that which is obtained from 1 with the aid of P. Let this be l_2. We take as the third element that which is obtained from l_2 with the aid of P, and so on, until finally we arrive at the element which becomes 1 with the aid of P. This will be the last element of the first cycle. It may easily be seen that this cycle cannot contain identical elements. The cycle thus composed does not in general exhaust all n elements. We take any one of the remaining elements as first element of a new cycle and form this as above, and so on.

We take as an example the permutation with $n = 6$:

$$\begin{pmatrix} 1, 2, 3, 4, 5, 6 \\ 3, 6, 4, 1, 2, 5 \end{pmatrix}.$$

By using the above method, we can write this as the product of the cycles

$$\begin{pmatrix} 1, 2, 3, 4, 5, 6 \\ 3, 6, 4, 1, 2, 5 \end{pmatrix} = (1, 3, 4,)\,(2, 6, 5),$$

the order of the factors on the right being of no consequence.

The product of two transpositions is readily expressible as a product of third degree cycles. If the second degree cycles (transpositions) have no common elements, it may easily be shown that

$$(l_3, l_4)\,(l_1, l_2) = (l_1, l_3, l_4)\,(l_1, l_2, l_4),$$

whereas with common elements:

$$(l_1, l_3)\,(l_1, l_2) = (l_1, l_2, l_3)\,.$$

Thus every permutation of an alternating group can be written as a product of third degree cycles.

We also note that the numbers in the first row of a permutation can be written in any order. The only thing that matters is that each of these should have below it the number which it becomes as a result of the given permutation. For instance, the following are forms of the same permutation:

$$\begin{pmatrix} 1, 2, 3, 4, 5 \\ 3, 2, 5, 1, 4 \end{pmatrix} = \begin{pmatrix} 3, 1, 5, 4, 2 \\ 5, 3, 4, 1, 2 \end{pmatrix}.$$

Given the permutation

$$P = \begin{pmatrix} a_1, a_2, \ldots, a_n \\ b_1, b_2, \ldots, b_n \end{pmatrix},$$

we can obviously write the inverse permutation as

$$P^{-1} = \begin{pmatrix} b_1, & b_2, & \dots, & b_n \\ a_1, & a_2, & \dots, & a_n \end{pmatrix}.$$

Suppose we have two permutations, the second being written in two ways:

$$P = \begin{pmatrix} 1, & 2, & \dots, & n \\ b_1, & c_2, & \dots, & c_n \end{pmatrix}; \quad Q = \begin{pmatrix} 1, & 2, & \dots, & n \\ d_1, & d_2 & \dots, & d_n \end{pmatrix} = \begin{pmatrix} c_1, & c_2, & \dots, & c_n \\ f_1, & f_2, & \dots, & f_n \end{pmatrix}.$$

We have

$$PQ^{-1} = \begin{pmatrix} 1, & 2, & \dots, & n \\ c_1, & c_2, & \dots, & c_n \end{pmatrix} \begin{pmatrix} d_1, & d_2, & \dots, & d_n \\ 1, & 2, & \dots, & n \end{pmatrix} = \begin{pmatrix} d_1, & d_2, & \dots, & d_n \\ c_1, & c_2, & \dots, & c_n \end{pmatrix}$$

and consequently

$$QPQ^{-1} = \begin{pmatrix} c_1, & c_2, & \dots, & c_n \\ f_1, & f_2 & \dots, & f_n \end{pmatrix} \begin{pmatrix} d_1, & d_2, & \dots, & d_n \\ c_1, & c_2, & \dots, & c_n \end{pmatrix} = \begin{pmatrix} d_1, & d_2, & \dots, & d_n \\ f_1, & f_2, & \dots, & f_n \end{pmatrix}.$$

The following rule follows from this equation: to obtain the permutation QPQ^{-1}, carry out the permutation Q in both rows of

$$P = \begin{pmatrix} 1, & 2, & \dots, & n \\ c_1, & c_2, & \dots, & c_n \end{pmatrix}.$$

56. Abstract groups. When defining a group, we can completely disregard the concrete interpretation of the operations, the set of which forms the group, and which we have previously taken to be linear transformations or permutations. We arrive in this way at an *abstract group*.

An abstract group is a set of symbols for which multiplication is defined, in the sense that there is a definite rule by which two elements P and Q (the same or different) of the set yield a third element, also belonging to the set, which is called their product and written QP. The following three conditions must be fulfilled here.

1. *Multiplication must obey the associative law*, i.e. $(RQ)P = R(QP)$, whence it follows that, in general, any number of factors in a product can be grouped together, without, of course, changing their order.

2. *There must be one and only one element E in our set which, when multiplied on either side by any other element, yields the same element* i.e.

$$EP = PE = P. \tag{32}$$

We shall call E the *identity or unit element*.

3. *For any element P of the set, there exists a further unique element Q of the set which satisfies the condition*

$$QP = PQ = E \qquad (Q = P^{-1}). \tag{33}$$

With $P = E$, (32) gives $EE = E$, i.e. the inverse of E is E by definition of inverse element ($E^{-1} = E$).

These conditions defining an abstract group can be put in a more compact form with more restricted requirements in which case the restricted requirements imply the remainder as necessary formal consequences; we shall not dwell on this, however. On the whole we shall confine ourselves to the elementary basic facts regarding abstract groups. A detailed treatment of the theory of groups provides enough material to fill a separate volume. Our aim is simply to familiarize the reader with basic concepts and facilitate his reading of the literature of physics, where the group concept and the fundamental properties of groups are frequently utilized. Below, we shall occasionally write I instead of E. The element Q defined by equations (33) is called the *inverse* of P and is written P^{-1}. Equation (28) is clearly valid, since it follows from (33) that P is the inverse of Q.

Having established the concept of abstract group, we explain next some fresh concepts and also prove some properties of abstract groups. We first of all notice that the number of elements in a group can be either finite or infinite, as we saw above. Suppose we take the product of elements of a group

$$RQP.$$

This is also an element of the group. The inverse is obtained exactly as in the linear transformation group, i.e. it is

$$(RQP)^{-1} = P^{-1}Q^{-1}R^{-1}.$$

This is easily seen by carrying out the multiplication and using the associative law. Given an element P of the group, its positive integral powers

$$P^0 = I,\ P^1,\ P^2 \ldots$$

are likewise elements of the group. If there exists a positive integer m such that $P^m = I$, the element is said to be of *finite order*, the *order of an element* being the least positive number m for which $P^m = I$. There can be no identical elements among

$$I,\ P,\ P^2,\ \ldots,\ P^{m-1}.$$

For it follows at once from the condition $P^k = P^l$ ($k < l$) that $P^{l-k} = I$. All the elements of a finite group are clearly of finite order.

Let us write P_a for elements of the group. If the group is finite, a can be assumed to take a finite number of positive integral values. If the group is infinite, it can take all integral values [52], it can vary continuously, or it can even be equivalent to several subscripts which vary continuously. Let U be a fixed element of the group. We form all the possible products UP_a. We can easily show that, as the subscript a varies, the product again gives us all the elements of the group, without repetition.

For on multiplying on the left by U^{-1}, the equation

$$UP_{a_1} = UP_{a_2}$$

gives us at once $P_{a_1} = P_{a_2}$, i.e. the products UP_a must be different for different a. To show that the product can become any element of the group, we take $UP_a = P_{a_0}$, which is eqivalent to $P_a = U^{-1} P_{a_0}$, i.e. UP_a in fact gives us the element P_{a_0} when the factor P_a is equal to the element $U^{-1} P_{a_0}$ of the group. We should have the same result if the fixed element U were written on the right instead of the left. We thus arrive at the following: *if P_a varies over all the elements of the group and U is a fixed element, the product UP_a (or $P_a U$) likewise varies over all the elements, without repetition.*

We take as a particular example the group consisting of six elements (a sixth order group), the elements being denoted by

$$E,\ A,\ B,\ C,\ D,\ F.$$

We define the multiplication rule with the aid of the following table:

	E	A	B	C	D	F
E	E	A	B	C	D	F
A	A	E	D	F	B	C
B	B	F	E	D	C	A
C	C	D	F	E	A	B
D	D	C	A	B	F	E
F	F	B	C	A	E	D

(34)

The table must be used as follows. Suppose we want to find the product DB, we look for B in the first row and D in the first column then find A at the intersection of the corresponding column and row, so that the product DB is A. All the conditions laid down in the definition of abstract group may readily be seen to be satisfied here, the role of the identity element being played by E.

We have met with concrete interpretations of the abstract group concept in previous examples. In one case the role of element was played by linear transformations (their matrices) and multiplication of two elements amounted to successive application of two transformations, i.e. to multiplication of the corresponding matrices. In another case permutations played the part of elements and multiplication of two elements amounted to successively carrying out two permutations. We shall now mention a further concrete interpretation of the elements of a group.

Let the elements be all the complex numbers and let multiplication of two elements amount to addition of the corresponding complex numbers. In this case the role of identity element is played by zero, whilst the inverse element to the complex number a is $(-a)$. Instead of the complex numbers, we could have taken as elements all the vectors $\mathbf{x}(x_1, x_2, \ldots, x_n)$ of complex n-dimensional space R_n and defined the multiplication of elements as addition of the corresponding vectors. Here the null vector plays the part of identity element. We can say alternatively that the group elements are the vectors of R_n whilst the group operation is vector addition. We notice that in the last two examples the result of multiplying two elements of the group is independent of the order of the factors, i.e. as we say, any two elements of the group commute. A group of this type is called *Abelian* [cf. 45]. The simplest example of Abelian group is the *cyclic group* which consists of the identity element E and powers of an element P. If m is the least positive integer for which $P^m = E$, the cyclic group has m elements: $E, P, P^2, \ldots, P^{m-1}$. If there is no such positive integer m, the cyclic group is infinite: $E, P, P^2, \ldots$

57. Subgroups. Suppose a set H of only part of the elements of a given group G likewise forms a group, the above definition of multiplication being preserved. In this case the group H is said to be a subgroup of G. The set consisting simply of the identity (unit) element of G clearly always forms a subgroup. This is a trivial case that we shall overlook when speaking in future of subgroups.

We write H_a for elements of the subgroup H, and let G_1 be a given element, not belonging to H, of the total group G. As seen above, the products $G_1 H_a$ give various elements of G; these elements do not belong to H, since otherwise we should have for certain values a_1 and a_2 of the subscript a: $G_1 H_{a_1} = H_{a_2}$, whence $G_1 = H_{a_2} H_{a_1}^{-1}$, i.e. G_1 must belong to H, which contradicts our assumption. Now let G_1

and G_2 be two different elements of G not belonging to the subgroup H. We show that the sets of elements $G_1 H_a$ and $G_2 H_a$ either have no common elements at all or else coincide, i.e. consist of the same elements. For suppose we have $G_2 H_{a_2} = G_1 H_{a_1}$ for certain values of a; it follows that $G_2 = G_1 H_{a_1} H^{-1}{}_{a_2} = G_1 H_{a_3}$, i.e. G_2 belongs to the set of elements $G_1 H_a$, and similarly G_1 belongs to the set of $G_2 H_a$. Hence the products $G_1 H_a$ and $G_2 H_a$ define the same set of elements.

We take all the elements H_a of the subgroup H. These do not exhaust the elements of G. We take some element G_1 not belonging to H and form all the products $G_1 H_a$ which, as we have seen, all differ from each other and from the H_a.

It can happen that the H_a and $G_1 H_a$ do not exhaust the whole group. We take an element G_2 not belonging to the H_a or $G_1 H_a$ and form all the products $G_2 H_a$. As we have seen, the elements $G_2 H_a$ all differ from each other and from the H_a and $G_1 H_a$. If the elements H_a, $G_1 H_a$ and $G_2 H_a$ do not exhaust G, we take an element G_3 not belonging to any of the above three sets and form the products $G_3 H_a$. Hence we get further elements of the group, and so on. Suppose we exhaust the elements of G by means of a finite number of such operations. Let $(m - 1)$ elements G_k be required for this. All the elements of G are now represented as follows:

$$H_a,\ G_1 H_a,\ G_2 H_a,\ \ldots,\ G_{m-1} H_a, \tag{35}$$

where the subscript a varies over values corresponding to the subgroup H. If we set $G_k' = G_k H_{a_0}$, where a_0 is fixed in some manner, the set of elements $G_k' H_a$ will coincide, as shown above, with the set $G_k H_a$. In other words, in every set $G_k H_a$ $(G_0 = I)$ any element of the set can play the role of G_k. Hence it follows at once that, for any given subgroup H_a, the division of the elements of group G into sets of type (35) is fully defined. The $G_k H_a$ are called *cosets with respect to the subgroup* H_a.

In the case of (35), H is said to be a *subgroup of finite index*, and is in fact a *subgroup of index* m. If G is a finite group, the index of the subgroup H is clearly equal to the quotient of the order of G and the order of H, *the order of a finite group being defined as the number of elements contained in it.* We notice that only the first of sets (35) forms a subgroup. The remaining sets $G_k H_a$ do not contain the identity element and therefore cannot form a subgroup.

We formed scheme (35) by multiplying the elements H_a of the subgroup H on the left by elements G_k of group G. We could multiply

on the right; changing the notation from G_k to G'_k, we arrive at the following new representation of the elements of G:

$$H_\alpha,\ H_\alpha G'_1,\ H_\alpha G'_2,\ \ldots,\ H_\alpha G'_{m-1}, \tag{36}$$

where it will be shown that the index m of the subgroup remains unchanged. The sets of elements $G_k H_\alpha$ are sometimes called *left cosets* whilst the $H_\alpha G'_k$ are right cosets.

We first of all observe that, if α varies over all the values corresponding to the subgroup H, the elements H_α^{-1} give all the elements of H. This follows directly from the fact that the inverse of a given element of H also belongs to H. We now turn to the proof that the indices for right and left cosets are the same. We take any two different sets $G_p H_\alpha$ and $G_q H_\alpha$ $(p \neq q)$ of (35), where we can suppose for the first of sets (35) that say $G_p = E$. We take the inverse elements:

$$(G_p H_\alpha)^{-1} = H_\alpha^{-1} G_p^{-1} \quad \text{and} \quad (G_q H_\alpha)^{-1} = H_\alpha^{-1} G_q^{-1}.$$

Bearing in mind the remarks made above, we can re-write these sets of elements as $H_\alpha G_p^{-1}$ and $H_\alpha G_q^{-1}$. They are easily seen to have no common elements. For suppose we had

$$H_{\alpha_1} G_p^{-1} = H_{\alpha_2} G_q^{-1},$$

it would follow that

$$G_p^{-1} G_q = H_{\alpha_1}^{-1} H_{\alpha_2} = H_{\alpha_3}, \quad \text{or} \quad G_q = G_p H_{\alpha_3},$$

and G_q would belong to the set $G_p H_\alpha$, which is impossible. Hence the sets

$$H_\alpha,\ H_\alpha G_1^{-1},\ H_\alpha G_2^{-1},\ \ldots,\ H_\alpha G_{m-1}^{-1}$$

are seen to be right cosets, so that we can simply take $G'_s = G_s^{-1}$ in (36).

We consider some examples of subgroups. Let G be the set of real orthogonal transformations in three variables and H the set of real orthogonal transformations in three variables with $(+1)$ determinant. Every real orthogonal transformation is either a rotation, i.e. belongs to H, or is the product of a rotation and the symmetric reflection relative to the origin given by

$$x' = -x; \quad y' = -y; \quad z' = -z. \quad (S) \tag{37}$$

The present group G can be represented by the scheme

$$H_\alpha,\ SH_\alpha \tag{38}$$

or

$$H_\alpha,\ H_\alpha S, \tag{39}$$

where H_a denotes the set of all elements of the group H. Here, H_a is a subgroup of index 2

Let G be the symmetric group of permutations of n elements, and H the alternating group made up of even permutations. Further, let S be any given odd permutation, say the permutation consisting of the single cycle (1, 2), i.e. amounting to transposition of the elements 1 and 2. It is clear that here also we can represent G by scheme (38) or (39). In both cases, multiplication on the left leads to the same result as multiplication on the right.

Here, the alternating group is a subgroup of index two of the symmetric group.

We also consider the finite group of the regular octahedron that we discussed above. Let l be the axis passing through a given vertex A of the octahedron. Let S_0, S_1, S_2, S_3 be the rotations about this axis by angles $0, \pi/2, \pi$, and $3\pi/2$. These rotations form a subgroup of the total group of rotations of the octahedron. Let T_k denote the rotations displacing A to the remaining five vertices ($k = 1, 2, 3, 4, 5$). We can write the complete octahedral group as

$$S_a,\ T_1 S_a,\ T_2 S_a,\ T_3 S_a,\ T_4 S_a,\ T_5 S_a,$$

i.e. S_a is a subgroup of index six.

Let G_s, G_s^{-1} be elements of group G ($s = 1, 2, \ldots, k$). We consider the set of all elements of G expressible as products of G_s, G_s^{-1} ($s = 1, 2, \ldots, k$).

This set clearly forms a group which is a subgroup of G or else coincides with G.

This subgroup is said to be generated by the given set of elements G_s, G_s^{-1} ($s = 1, 2, \ldots, k$).

58. Classes and normal subgroups. Let U and V be elements of a group. The element $W = VUV^{-1}$ is said to be *conjugate to* U. It is easily seen that, conversely, U is conjugate to W. For $U = V^{-1}WV$. Two elements U_1 and U_2, conjugate to a third element W:

$$U_1 = V_1 W V_1^{-1};\quad U_2 = V_2 W V_2^{-1},$$

are also conjugate to each other:

$$U_2 = V_2 V_1^{-1} U_1 (V_2 V_1^{-1})^{-1}.$$

The set of all mutually conjugate elements of a group forms what is known as a *class of the group*. A class is fully defined by one of its

elements U. For, given U, we get the entire class by the expression $G_a UG_a^{-1}$, where G_a varies over all the elements of the group. We can thus divide the total group into classes. Bearing in mind the basic property of the identity element, given in [56], we have

$$G_a I G_a^{-1} = I,$$

i.e. the identity element itself forms a class.

If the element U is of order m, i.e. m is the least positive integer for which $U^m = I$, every conjugate element $G_a UG_a^{-1}$ has the same order m, as follows at once from the equations

$$(G_a U G_a^{-1})^m = G_a U^m G_a^{-1} = I.$$

In other words, all elements of the same class have the same order.

We remark that when G_a runs over all the elements of group G, the product $G_a UG_a^{-1}$ can give the elements of the class more than once. For instance, if $U = I$, the product always gives I, as we have seen.

We again take as an example the octahedral rotation group. Let U be the rotation by $\pi/2$ about an axis A_pA_q of the octahedron. If the rotation T_k, belonging to the group, transforms the axis l to the axis l_1, the vertex A_p being transformed to A_r and A_q to A_s, the group element $T_k UT_k^{-1}$ yields a rotation by $\pi/2$ about the axis $A_r A_s$. If, for instance, T_k transforms A_p to A_q, the product here yields a rotation by $\pi/2$ about the axis A_qA_p or, in other words, a rotation by $3\pi/2$ about A_pA_q. If T_k transforms $A_p A_q$ into itself, i.e. if the rotation is about this axis, the product $T_kUT_k^{-1}$ coincides with U. Thus in the present case, the class of elements conjugate to U consists of the set of rotations by $\pi/2$ about axes of the octahedron.

Similarly, taking the group of rotations of three-dimensional space about the origin, we know that every group element U consists of a rotation by some angle φ about a certain axis. Here the class of elements conjugate to U is the set of rotations by the angle φ about all the possible axes passing through the origin.

We now discuss another important concept, that of a *normal subgroup*; this is closely connected with the idea of class. Let H be a subgroup of a group G, and G_1 a fixed element of G. We consider the set of group elements given by

$$G_1 H_a G_1^{-1}, \tag{40}$$

where H_a denotes a variable element of the subgroup H, in other

words, H_a runs over all the elements of H. Products (40) may easily be seen also to form a subgroup. For if we take say the product of two elements belonging to the set (40), it also belongs to this set:

$$(G_1 H_{a_2} G_1^{-1})(G_1 H_{a_1} G_1^{-1}) = G_1 H_{a_2} H_{a_1} G_1^{-1} = G_1 G_{a_3} G_1^{-1},$$

and the other conditions for group formation are similarly fulfilled.

Subgroup (40) is said to be conjugate to subgroup H; if G_1 belongs to H, (40) also consists of elements belonging to H, and as is easily seen, simply coincides with H.

Every element H_{a_0} of subgroup H can in this case be obtained from (40), if we take

$$H_a = G_1^{-1} H_{a_0} G_1.$$

If the element G_1 does not belong to subgroup H, subgroup (40) can be different from H.

Subgroup H is called a *normal subgroup* of the total group G if, for any choice of element G_1 of the total group G, subgroup (40) coincides with H. We now discuss some new concepts connected with normal subgroups, examples of which will be given later.

Let subgroup H be a normal subgroup of the total group G. We shall simplify the writing by assuming this subgroup to have a finite index m. In this case, all elements of group G can be represented by

$$H_a,\ G_1 H_a,\ G_2 H_a,\ \ldots,\ G_{m-1} H_a, \tag{41}$$

where, as usual, H_a is a variable element of H. Since H is a normal subgroup, the set of elements $G_k H_a G_k^{-1}$ coincides with the set of elements H_a, i.e. the set of elements $G_k H_a$ is the same as the set of elements $H_a G_k$.

Hence, if *H is a normal subgroup*, the division of the group elements into conjugate sets in accordance with (41) is the same as the division into conjugate sets according to the scheme

$$H_a,\ H_a G_1,\ H_a G_2,\ \ldots,\ H_a G_{m-1}. \tag{42}$$

In other words, in this case *right cosets are the same as left cosets*.

Given any element H_{a_0} of the normal subgroup, the element $G_0 H_{a_0} G_0^{-1}$, for any choice of G_0 from G, also belongs to the normal subgroup, i.e. if an element belongs to the normal subgroup, the entire class in which the element appears in the basic group also belongs to the normal subgroup. We can easily prove the converse: a subgroup

is a normal subgroup if, on containing an element, it contains the entire class to which the element belongs in the basic group.

We now return to the cosets of (41) or (42), where the elements H_a make up the normal subgroup H. We consider the products $G_l H_a G_k H_{a'}$ of elements of a coset $G_l H_a$ with elements of the coset $G_k H_{a'}$.

We can write the set of these products as

$$G_l (H_a G_k) H_{a'},$$
$$G_l G_k H_a H_{a'}.$$

The elements H_a and $H_{a'}$ belong to the normal subgroup H, and the same can be said of their product. Hence we can write the above products as

$$G_l G_k H_a.$$

All elements of this type are included in the same coset of (41), viz., the set to which the element $G_l G_k$ belongs. It is also easily shown that in this way we get all the elements of this coset. In short, if a subgroup is a normal subgroup, multiplication of one conjugate set by another likewise gives a coset. We shall look on the cosets as new elements, the first set in scheme (41), (38) being taken as the identity element. The above result regarding the multiplication of cosets gives us a multiplication rule for these new elements that we have introduced. We propose to the reader the simple proof that this multiplication rule satisfies all the requirements for group formation, i.e. given this rule, our new elements themselves form a group, in which the first coset of the scheme plays the part of identity element. This new group, the order of which is equal to the index of the normal subgroup H, is said to be *complementary to H* or is called the *factor group relative to H*.

Every group G has two trivial normal subgroups: the identity element by itself, and the entire group G.

We shall in future assume a non-trivial case when speaking of normal subgroups. It may happen that a group has no normal subgroup.

The group is said to be *simple* in this case.

59. Examples. 1. We take the group G of real orthogonal transformations in three-dimensional space. Let H be the subgroup of rotations, i.e. the set of orthogonal transformations with $(+1)$ determinant. Further, let S be the symmetrical reflection about the origin defined by (37). If H_a is a variable

element of H, the total group G can be represented as

$$H_a, \; SH_a \quad \text{or} \quad H_a, \; H_a S. \tag{43}$$

If G_1 is any transformation of G, $G_1 H_a G_1^{-1}$ has a $(+1)$ determinant, i.e. belongs to H, and H is a normal subgroup of index two. We consider the complementary group to H. The identity element E of this group corresponds to the first of sets (43). The product of two elements of the second set, i.e. of two orthogonal transformations with (-1) determinant, yields an orthogonal transformation with $(+1)$ determinant which belongs to the first set. If K is the element corresponding to the second set, it follows from what has been said that $K^2 = E$. Thus the complementary group to H consists of the two elements E and K, and $K^2 = E$, i.e. it is a cyclic group of order two. This is true in general for normal subgroups of index two.

2. For the symmetric group of permutations, the alternating group is a normal subgroup of index two.

We write down the elements of the symmetric group with three elements and denote each by a single letter, using the notation of [55]:

$$E; \quad A = (2, 3); \quad B = (1, 2); \quad C = (1, 3); \quad D = (1, 3, 2); \quad F = (1, 2, 3).$$

The alternating group consisting of permutations E, D, E, is a third order cyclic group ($F = D^2$ and $D = F^2$), where $D^3 = F^3 = E$. The total symmetric group consists of three classes: I E; II A, B, and C; III D and F.

The alternating group also consists of three classes: I E; II D; III F. It is easily verified that the multiplication rule for elements of the symmetric group in question is that defined by table (34) of [56].

The alternating group with $n = 4$ contains twelve elements which are distributed in four classes;

$$\text{I } E; \quad \text{II } A_1 = (1, 2)(3, 4); \quad A_2 = (1, 3)(2, 4); \quad A_3 = (1, 4)(2, 3);$$
$$\text{III } B_1 = (1, 2, 3); \quad B_2 = (2, 1, 4); \quad B_3 = (3, 4, 1); \quad B_4 = (4, 3, 2);$$
$$\text{IV } C_1 = (1, 2, 4); \quad C_2 = (2, 1, 3); \quad C_3 = (3, 4, 2); \quad C_4 = (4, 3, 1).$$

The second class contains three second order elements, whilst the third and fourth each have four third order elements. The product of two elements of the second class may easily be seen to yield an element again of the second class, and since all second order elements fall into the second class, we can say that these three elements form, together with the identity element, a normal subgroup of the alternating group. Its order is four, whilst the index is three. It is easily verified that elements B_i of the third class fall into one of the cosets of elements with respect to this normal subgroup, whilst the elements C_i are in the other coset. It may further readily be seen that the product of two third class elements yields a fourth class element, whilst the product of two fourth class elements yields a third class element. The identity element E in the complementary group corresponds to the above normal subgroup. Let A and B be the two other elements of the complementary group. It follows immediately from what was said above that $A^2 = B$ and $B^2 = A$, and it is at once clear that the complementary group consisting of the elements E, A and A^2, where $A^3 = E$, is a cyclic group of the third order.

We observe that the elements E, (1, 2, 3) and (2, 1, 3) of the basic alternating group form a cyclic subgroup of the third order, though this subgroup is not an elementary divisor.

On enumerating the vertices of a tetrahedron in any order, it is easily verified directly that the above alternating group with $n = 4$ corresponds to the rotations that displace the tetrahedron into itself. Every permutation defines a passage from one vertex to another. Rotations of $2\pi/3$ about an axis of the tetahedron correspond to permutations of the third class, whilst opposite rotations by the same angle about the same axes correspond to permutations of the fourth class. For instance, rotations about axes passing through vertex number four correspond to the permutations (1, 2, 3) and (2, 1, 3). Rotations such that no vertex remains unchanged correspond to the permutations of the second class.

The alternating group can be shown to be simple with $n > 4$.

3. If H is any subgroup of an Abelian group G, we have $G_1 H_a = H_a G_1$ for any choice of elements H_a of H and G_1 of G, i.e. $G_1 H_a G_1^{-1} = H_a$, whence it is at once clear that H is a normal subgroup, i.e. every subgroup of an Abelian group is a normal subgroup. We take as an example the group G of vector addition in R_n which was mentioned in [49].

We take for the subgroup H the vectors belonging to some subspace L_k of R_n $(0 < k < n)$. The cosets are obtained by associating with any vector $\mathbf{x}$ of R_n all the vectors of the subspace L_k.

If $\mathbf{x}$ belongs to L_k, the coset is the same as the subgroup H. We introduce fundamental vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$ into L_k and fundamental vectors $\mathbf{x}^{(k+1)}, \ldots, \mathbf{x}^{(n)}$ into the complementary subspace M_{n-k}. By what has been said above, every coset of elements consists of the vectors:

$$c_1 \mathbf{x}^{(1)} + c_2 \mathbf{x}^{(2)} + \ldots + c_k \mathbf{x}^{(k)} + c_{k+1} \mathbf{x}^{(k+1)} + \ldots + c_n \mathbf{x}^{(n)},$$

where $c_{k+1}, \ldots, c_n$ have fixed values and $c_1, c_2, \ldots, c_k$ are arbitrary.

We can thus associate with every coset a definite vector of M_{n-k}, and conversely, a definite coset corresponds to every vector of M_{n-k}. Corresponding to addition of two vectors of any two cosets we have the addition of the corresponding vectors of M_{n-k}. In other words, in the case of the above group operation (vector addition), we can regard the vectors of M_{n-k} as elements of the complementary group.

In this example, the order of the normal subgroup H and its index are infinite.

60. Isomorphic and homomorphic groups. Two groups A and B are said to be *isomorphic* if a one-to-one correspondence can be established between their elements, i.e. to any element of A there corresponds one element of B, and vice versa, the correspondence being such that the product of any two elements of A corresponds to the product of the corresponding elements of B. If A and B are isomorphic abstract groups, they have precisely the same structure, i.e. they have no essential differences.

We now turn to a new concept that represents a generalization of group isomorphism. Group B is said to be *homomorphic to group A* if to every element of A there corresponds a definite element of B, and to every element of B at least one element of A, the correspondence being such that to the product of two elements of A there corresponds the product of the corresponding elements of B. The present case differs from that of isomorphism in that the correspondence need not be one-to-one both ways, i.e. the same element of B can correspond to several different elements of A. If B is homomorphic to A, and one definite element of A corresponds to each element of B, the groups are also isomorphic. We observe moreover that if the elements B_1 and B_2 of B correspond to A_1 and A_2 of A, the element $B_2 B_1$ corresponds by definition to the element $A_2 A_1$ of A.

Let A_0 be the identity (unit) element of A and B_0 the corresponding element of B. We can easily show that B_0 is the identity element of B. For, we have for any A_1 of A:

$$A_0 A_1 = A_1 A_0 = A_1,$$

which leads to the equation for the corresponding elements of B:

$$B_0 B_1 = B_1 B_0 = B_1,$$

where B_1 can be assumed to be any element of B by the definition of homomorphism. The last equation shows that B_0 is the identity element of B. Thus the identity element of B corresponds to the identity element of A in isomorphic and homomorphic groups. We now take A_1 and its inverse A^{-1} of A, and let B_1 and B_2 be the corresponding elements of B. The equation $A_1 A_1^{-1} = A_1^{-1} A_1 = A_0$, where A_0 is the identity element of A, gives by definition of homomorphic groups, $B_1 B_2 = B_2 B_1 = B_0$, where B_0 is the identity element by the above; thus $B_2 = B_1^{-1}$, i.e. inverse elements of B correspond to inverse elements of A.

Suppose that our groups are homomorphic, but not isomorphic. We consider the set of elements C_a in group A to which the identity element B_0 of B corresponds. If B_0 corresponds to C_a, $B_0^{-1} = B_0$ corresponds to C_a^{-1} by the above and we have $B_0 B_0 = B_0$ corresponding to every product $C_{a_2} C_{a_1}$, i.e. the set of elements C_a of A to which the identity element of B corresponds forms a subgroup C of A.

We show that this subgroup is a normal subgroup. Let A_1 be any given element of A and B_1 the corresponding element of B. To every element of the form $A_1 C_a A_1^{-1}$ there corresponds the element $B_1 B_0 B_1^{-1}$

of B, this latter being B_0 in view of the basic properties of identity elements; thus every element of the form $A_1 C_a A_1^{-1}$ is one of the elements C_a, i.e. it belongs to the subgroup C, or in other words, C is a normal subgroup. We now consider the decomposition of A into cosets as follows:

$$C_a, \; A_1 C_a, \; A_2 C_a, \ldots \tag{44}$$

Let B_k be the element corresponding to A_k. We take two elements $A_k C_{a_1}$ and $A_k C_{a_2}$ belonging to the same conjugate set. The corresponding elements to these are $B_k B_0$ and $B_k B_0$, i.e. the same element B_k of B.

Elements B_k and B_l correspond to different cosets $A_k C_a$ and $A_l C_a$. We show that B_k and B_l are different. If they were the same, the identity element B_0 of B would correspond to the element $A_k^{-1} A_l$, i.e. this latter element would be one of the C_a and we should have $A_k^{-1} A_l = C_{a_0}$, i.e. $A_l = A_k C_{a_0}$, which contradicts scheme (44). Hence, *if group B is homomorphic to group A, the set of elements C of A corresponding to the identity element of B form a normal subgroup, and each coset to this normal divisor is a set of all the elements of A corresponding to a given element of B.* It follows directly from the definition of homomorphic groups moreover, that we have, corresponding to the product of any two elements of different (or the same) cosets, the product of the elements of B corresponding to these sets, i.e. more briefly, a definite element of B corresponds to each coset of A, different elements of B correspond to different cosets, and the correspondence establishes an isomorphism between group B and the group in A complementary to C_a.

We again take as an example the group of real orthogonal transformations in three-dimensional space; we associate with each transformation a number equal to the corresponding determinant, and define multiplication in the domain of these numbers in the usual way, i.e. as numerical multiplication. Our group is now homomorphic to the group consisting of the two elements $(+1)$ and (-1), multiplication being defined in the usual numerical way for these two elements. The role of identity element is played by $(+1)$. The normal subgroup in this example consists of the group of rotations.

If group B is homomorphic but not isomorphic to group A, the set of elements of A to which the identity element of B corresponds is generally known as the *kernel of the homomorphism.* We have seen that the kernel of the homomorphism is a normal subgroup of group A.

61. Examples. 1. We take the group G of real orthogonal transformations in three-dimensional space and associate with each the number equal to the determinant of the transformation, the group operation for these numbers being defined as ordinary multiplication. The group G', consisting of $(+1)$ and (-1) with ordinary multiplication of these numbers, is now homomorphic to group G. The identity element $(+1)$ of G' corresponds to rotations of three-dimensional space of G. These rotations form a normal divisor, whilst the complementary group is the cyclic group of order two [59].

2. We take an equilateral triangle in the xy plane with vertices

$$(1,\ 0);\ (\cos 120°,\ \sin 120°);\ (\cos 240°,\ \sin 240°)$$

and form the group G consisting of rotations of the plane about the origin by the angles 0°, 120°, and 240° which displace the triangle into itself, and of reflection in the x axis followed by rotation by the angles 0°, 120°, 240°. This is the dihedral group with $n = 3$.

We write down all the matrices corresponding to the group elements:

$$E = \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix};\quad A = \begin{Vmatrix} 1, & 0 \\ 0, & -1 \end{Vmatrix};\quad B = \begin{Vmatrix} -\frac{1}{2}, & \frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3}, & \frac{1}{2} \end{Vmatrix};$$

$$C = \begin{Vmatrix} -\frac{1}{2}, & -\frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3}, & \frac{1}{2} \end{Vmatrix};\quad D = \begin{Vmatrix} -\frac{1}{2}, & \frac{1}{2}\sqrt{3} \\ -\frac{1}{2}\sqrt{3}, & -\frac{1}{2} \end{Vmatrix};\quad F = \begin{Vmatrix} -\frac{1}{2}, & -\frac{1}{2}\sqrt{3} \\ \frac{1}{2}\sqrt{3}, & -\frac{1}{2} \end{Vmatrix}.$$

If we return to the multiplication scheme given by table (34) of [56], we see that this scheme in fact corresponds to multiplication of the matrices forming our group. We saw above [59] that this scheme of multiplication also corresponds to the symmetric group of permutations of three elements:

$$E;\ A = (2, 3);\quad B = (1, 2);\quad C = (1, 3);\quad D = (1, 3, 2);\quad F = (1, 2, 3)\,. \quad (45)$$

Thus if we take as corresponding elements the elements of these two groups denoted by the same letter, the two groups are isomorphic. The permutations of group (45) correspond to permutations of the vertices of the equilateral triangle, if theso are numbered correspondingly.

Following the same lines as our discussion of [59], the tetrahedral group is isomorphic to the alternating group with $n = 4$.

3. A general method can be shown for constructing groups of permutations homomorphic to a given group G. Let H be any subgroup of finite index n of group G. We write down the cosets of this:

$$H,\ HS_1,\ HS_2,\ \ldots,\ HS_{n-1}\,.$$

If we multiply each set on the right by some element S of G, the result is merely a permutation of the order of the sets, and we shall take it that this permutation in fact corresponds to the chosen element S of G. It may easily be shown that we obtain in this way a group G' of permutations homomorphic to group G.

The necessary and sufficient condition for the identity element of G' to correspond to the element S of G is that, on multiplying on the right by S, every coset becomes itself, i.e.

$$H_\alpha S = H_\beta \text{ and } H_\alpha S_k S = H_\beta S_k \qquad (k = 1, 2, \ldots, n-1),$$

where H_α is any element of H and H_β also belongs to H. These equations can be re-written as

$$S = H_\alpha^{-1} H_\beta; \quad S = (S_k^{-1} H_\alpha S_k)^{-1} (S_k^{-1} H_\beta S_k),$$

and it now follows that the necessary and sufficient condition for the identity element of G' to correspond to the element S is that S belongs simultaneously to H and to every conjugate subgroup $S_k^{-1} HS_k$.

If H is a normal subgroup of G, the above requirement amounts to the fact that S belongs to H, and group G' is in this case isomorphic to the complementary group. If H is the identity element alone, G is isomorphic to the group of permutations G' which is obtained if the elements of G:

$$E, S_1, S_2, \ldots, S_n$$

are multiplied on the right by any given element S of G which reduces to some permutation of the elements of G. We shall discuss in more detail later the construction of groups of linear transformations isomorphic to a given group.

62. Stereographic projections. Having now concluded the fundamental general theory of groups, we consider a particular example of group correspondence that has an important role in physics. We start with a preliminary description of *stereographic projections*, that give a definite correspondence rule between points of a sphere and of a plane.

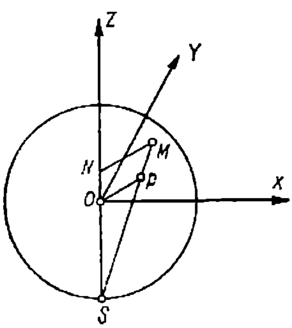

FIG. 3

We take xyz axes in three-dimensional space and a sphere C of unit radius with centre at the origin. Let S be the point of the sphere with coordinates $(0, 0, -1)$ and M a variable point on the sphere (Fig. 3). The straight line SM intersects the xy plane in a point P, so that we have a fully defined correspondence rule between points of the sphere C and points of the xy plane; the point corresponding to S $(0, 0, -1)$ on the sphere is the point at infinity on the plane. This point correspondence in fact gives the stereographic projection of the sphere on to the plane.

We now deduce the expressions for the stereographic projection. Let MN be the perpendicular from M to the z axis. We have from

similar triangles, bearing in mind that $SO = 1$:

$$NM = (1 + ON)\, OP\,. \tag{46}$$

Writing the coordinates of M as (x, y, z) and of P as (α, β), we have

$$NM = (1 + z)\, OP\,,$$

or, on projecting the parallel lines OP and NM on to the x and y axes:

$$x = (1 + z)\, \alpha; \quad y = (1 + z)\, \beta\,. \tag{47}$$

The equation $x^2 + y^2 + z^2 = 1$ gives us the quadratic equation for z:

$$(\alpha^2 + \beta^2)\,(1 + z)^2 + z^2 = 1\,,$$

and we obtain on solving this:

$$z = \frac{\pm 1 - (\alpha^2 + \beta^2)}{1 + (\alpha^2 + \beta^2)}\,.$$

But we must have $z > -1$ for all points (α, β) at a finite distance, and consequently we must have $(+1)$ in the above expression. On making use also of expressions (47), we get finally for (x, y, z) in terms of (α, β):

$$x = \frac{2\alpha}{1 + \alpha^2 + \beta^2}\,; \quad y = \frac{2\beta}{1 + \alpha^2 + \beta^2}\,; \quad z = \frac{1 - (\alpha^2 + \beta^2)}{1 + \alpha^2 + \beta^2}\,. \tag{48}$$

We introduce instead of the two real coordinates α, β on the plane the complex coordinate $\zeta = \alpha + i\beta$. On writing as usual $\bar{\zeta}$ for the complex conjugate of ζ, we can re-write the above expressions as

$$x + iy = \frac{2\bar{\zeta}}{1 + \zeta\bar{\zeta}}\,; \quad x - iy = \frac{2\zeta}{1 + \zeta\bar{\zeta}}\,; \quad z = \frac{1 - \zeta\bar{\zeta}}{1 + \zeta\bar{\zeta}}\,. \tag{49}$$

We write ζ as the ratio of two other complex numbers ξ and η:

$$\zeta = \frac{\eta}{\xi}\,. \tag{50}$$

Pairs of values of ξ and η, differing by a common factor, i.e. pairs of the form $k\xi$, $k\eta$ and ξ, η give the same ζ, i.e. the same point of the plane, whilst the pair $\eta \neq 0$, $\xi = 0$ gives the point at infinity. The complex numbers ξ and η are called homogeneous complex coordinates on the plane. Using (50) and separating into real and imaginary parts, we can re-write (49) as

$$x = \frac{\bar{\xi}\eta + \xi\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}}\,; \quad y = \frac{1}{i}\,\frac{\bar{\xi}\eta - \xi\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}}\,; \quad z = \frac{\xi\bar{\xi} - \eta\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}}\,. \tag{51}$$

These last formulae give us, for any complex ξ and η, real (x, y, z) satisfying the relationship

$$x^2 + y^2 + z^2 - 1 = 0 \,, \tag{51_1}$$

as is to be expected, since the point (x, y, z) lies on the unit sphere.

63. Unitary groups and groups of rotations. We now consider a unitary transformation of the variables (ξ, η):

$$\xi' = a\xi + b\eta, \quad \eta' = c\xi + d\eta, \tag{52}$$

where, since the transformation is unitary, we must have

$$\xi'\bar{\xi}' + \eta'\bar{\eta}' = \xi\bar{\xi} + \eta\bar{\eta} \,. \tag{53}$$

The new values (ξ', η') give us a new point on the sphere:

$$x' = \frac{\bar{\xi}'\eta' + \xi'\bar{\eta}'}{\xi'\bar{\xi}' + \eta'\bar{\eta}'}\,; \quad y' = \frac{1}{i}\,\frac{\bar{\xi}'\eta' - \xi'\bar{\eta}'}{\xi'\bar{\xi}' + \eta'\bar{\eta}'}\,; \quad z' = \frac{\xi'\bar{\xi}' - \eta'\bar{\eta}'}{\xi'\bar{\xi}' + \eta'\bar{\eta}'}\,. \tag{54}$$

We know that the determinant of unitary transformation (52) has unit modulus, so that it is given by a number of the form $e^{i\varphi}$. On multiplying all the coefficients of (52) by $e^{-i\varphi/2}$, we get a unitary transformation with determinant 1. But ξ' and η' are now also multiplied by $e^{-i\varphi/2}$. This extra factor has no effect at all on ζ. We can thus confine ourselves to unitary transformations (52) with the assumption of a unit determinant, i.e.

$$ad - bc = 1 \,. \tag{55}$$

Even with this restriction, two transformations, with coefficients of differing sign, give us values of ξ' and η' of differing sign, and we arrive at the same point ζ' under both transformations.

If we replace ξ' and η' in (54) by their expressions (52) and take condition (53) into account, we see on using (51) that the variables (x', y', z') are expressed as linear homogeneous polynomials in (x, y, z). By (53), the denominators are the same in (51) and (54), and variables (x, y, z) undergo the same linear transformation as the expressions

$$u = \bar{\xi}\eta + \xi\bar{\eta}\,; \qquad v = \frac{1}{i}(\bar{\xi}\eta - \xi\bar{\eta})\,; \qquad w = \xi\bar{\xi} - \eta\bar{\eta} \tag{56}$$

under unitary transformation (52). We establish later the exact form of this linear transformation.

We first of all establish the general form of unitary transformations (52) with unit determinant. The general conditions for unitary transformations yield [28]:

$$a\bar{c} + b\bar{d} = 0\,; \qquad c\bar{c} + d\bar{d} = 1\,.$$

On multiplying (55) by $\bar{c}$ and using the first condition written, we get

$$-\,bd\bar{d} - bc\bar{c} = \bar{c}\,,$$

whence we have by the second condition, $\bar{c} = -b$ or $c = -\bar{b}$, and we can show similarly that $d = \bar{a}$. Hence we can write all the unitary transformations with unit determinant as follows:

$$\left.\begin{aligned} \xi' &= a\xi + b\eta \\ \eta' &= -\,\bar{b}\xi + \bar{a}\eta, \end{aligned}\right\} \tag{57}$$

where a and b are any complex numbers satisfying the condition

$$a\bar{a} + b\bar{b} = 1\,. \tag{58}$$

We now write (56) with new variables

$$u' + iv' = 2\bar{\xi}'\eta'\,; \quad u' - iv' = 2\xi'\bar{\eta}'\,; \quad w' = \xi'\bar{\xi}' - \eta'\bar{\eta}',$$

or, using (57):

$$\begin{aligned} u' + iv' &= \bar{a}^2\, 2\bar{\xi}\eta - \bar{b}^2\, 2\xi\bar{\eta} - 2\bar{a}\bar{b}\,(\xi\bar{\xi} - \eta\bar{\eta}) \\ u' - iv' &= -\,b^2\, 2\bar{\xi}\eta + a^2\, 2\xi\bar{\eta} - 2ab\,(\xi\bar{\xi} - \eta\bar{\eta}) \\ w' &= \bar{a}b\, 2\bar{\xi}\eta + a\bar{b}\, 2\xi\bar{\eta} + (a\bar{a} - b\bar{b})\,(\xi\bar{\xi} - \eta\bar{\eta})\,. \end{aligned}$$

On making the substitutions

$$2\bar{\xi}\eta = u + iv\,; \quad 2\xi\bar{\eta} = u - iv\,; \quad \xi\bar{\xi} - \eta\bar{\eta} = w$$

and adding and subtracting the first two equations, we get expressions for (u', v', w') in terms of (u, v, w), or, what amounts to the same thing, expressions for (x', y', z') in terms of (x, y, z):

$$\left.\begin{aligned} x' &= \frac{1}{2}\left(a^2 + \bar{a}^2 - b^2 - \bar{b}^2\right)x\, + \\ &\quad + \frac{i}{2}\left(\bar{a}^2 + \bar{b}^2 - a^2 - b^2\right)y - \left(ab + \bar{a}\bar{b}\right)z \\ y' &= \frac{i}{2}\left(a^2 + \bar{b}^2 - \bar{a}^2 - b^2\right)x\, + \\ &\quad + \frac{1}{2}\left(a^2 + \bar{a}^2 + b^2 + \bar{b}^2\right)y + i\left(\bar{a}\bar{b} - ab\right)z \\ z' &= \left(a\bar{b} + \bar{a}b\right)x + i\left(a\bar{b} - \bar{a}b\right)y + \left(a\bar{a} - b\bar{b}\right)z\,. \end{aligned}\right\} \tag{59}$$

For every unitary transformation (57) there is a corresponding transformation of the xy plane, and this in turn, in view of the correspondence set up by the stereographic projection, gives a transformation of the sphere.

Accordingly, (59) is a real transformation by virtue of which the equation

$$x^2 + y^2 + z^2 = 1$$

becomes

$$x'^2 + y'^2 + z'^2 = 1.$$

But the linear homogeneous transformation (59) does not change the constant term 1, and consequently it must leave unchanged the left-hand side of the equation, i.e.

$$x'^2 + y'^2 + z'^2 = x^2 + y^2 + z^2.$$

All these facts can be obtained directly from the form of (59). Having shown that expressions (59) yield a real orthogonal transformation in three variables, we now show that the determinant of the transformation is always equal to $(+1)$. The determinant is a continuous function of the real and imaginary parts of the complex variables a and b, which must satisfy relationship (58). But the determinant can only have a value of $(+1)$ or (-1), and in view of the continuity the value must be either always $(+1)$ or always (-1). But with $a = 1$ and $b = 0$ expressions (59) give us the identity transformation with $(+1)$ determinant, i.e. the determinant of (59) is in fact always $(+1)$. Linear transformations (59) thus represent a rotation of space about the origin.

We now show that every rotation of space can be written in the form (59). If we set

$$a = e^{-\frac{i}{2}\varphi}; \quad \bar{a} = e^{\frac{i}{2}\varphi}; \quad b = \bar{b} = 0,$$

i.e. we take the unitary transformation matrix

$$A_\varphi = \begin{Vmatrix} e^{-\frac{i}{2}\varphi}, & 0 \\ 0, & e^{\frac{i}{2}\varphi} \end{Vmatrix}, \tag{60}$$

expressions (59) give us

$$\left.\begin{aligned} x' &= x\cos\varphi - y\sin\varphi, \\ y' &= x\sin\varphi + y\cos\varphi, \\ z' &= z, \end{aligned}\right\} \tag{61}$$

i.e. we get a rotation by the angle φ about the z axis.

If we now take

$$a = \bar{a} = \cos\frac{\varphi}{2}; \quad b = -i\sin\frac{\varphi}{2}; \quad \bar{b} = i\sin\frac{\varphi}{2},$$

i.e. we define the matrix of the unitary transformation as follows:

$$B_\varphi = \begin{Vmatrix} \cos\frac{\varphi}{2}, & -i\sin\frac{\varphi}{2} \\ -i\sin\frac{\varphi}{2}, & \cos\frac{\varphi}{2} \end{Vmatrix}, \tag{62}$$

(59) gives us

$$\left.\begin{aligned} x' &= x, \\ y' &= y\cos\varphi - z\sin\varphi, \\ z' &= y\sin\varphi + z\cos\varphi. \end{aligned}\right\} \tag{63}$$

This is a rotation by the angle φ about the x axis.

But as we know from [20], every rotation with Eulerian angles $\{\alpha, \beta, \gamma\}$ can be obtained as a result of rotation by the angle α about the z axis, followed by a rotation of β about the new x axis, followed by a rotation of γ about the new z axis.

If we write Z_φ for the third order matrix corresponding to transformation (61), and X_φ for the matrix of transformation (63), a rotation about the z axis by the angle α will be accomplished by the matrix Z_α, and the new x axis will now be obtained from the previous one with the aid of this matrix. A rotation of β about the new x axis will be accomplished, as may readily be seen, with the aid of $Z_\alpha X_\beta Z^{-1}_\alpha$, and these first two rotations are obtained by

$$Z_\alpha X_\beta Z_\alpha^{-1} Z_\alpha = Z_\alpha X_\beta.$$

As above, a rotation of γ about the new z axis is obtained by

$$(Z_\alpha X_\beta)\, Z_\gamma\, (Z_\alpha X_\beta)^{-1},$$

and finally the rotation $\{\alpha, \beta, \gamma\}$ is produced by

$$(Z_\alpha X_\beta)\, Z_\gamma\, (Z_\alpha X_\beta)^{-1} (Z_\alpha X_\beta)$$

or

$$Z_\alpha X_\beta Z_\gamma. \tag{64}$$

In the above arguments we have used the obvious fact that, if Z_φ is the matrix giving a rotation of φ about the l axis, passing through the origin, and M is a matrix transforming l to l_1, a rotation of φ about l_1 is given by the similar matrix

$$MZ_\varphi M^{-1}.$$

We now remark that, if A_1 and A_2 are two unitary transformations (57), corresponding to which we have orthogonal transformations (59) U_1 and U_2, the product $A_2 A_1$ will clearly correspond to $U_2 U_1$. Thus by (64), the rotation $\{\alpha, \beta, \gamma\}$ will be achieved by the unitary matrix consisting of the product of the three unitary matrices

$$\left\| \begin{matrix} e^{-i\frac{\alpha}{2}}, & 0 \\ 0, & e^{i\frac{\alpha}{2}} \end{matrix} \right\| \cdot \left\| \begin{matrix} \cos\frac{\beta}{2}, & -i\sin\frac{\beta}{2} \\ -i\sin\frac{\beta}{2}, & \cos\frac{\beta}{2} \end{matrix} \right\| \cdot \left\| \begin{matrix} e^{-i\frac{\gamma}{2}}, & 0 \\ 0, & e^{i\frac{\gamma}{2}} \end{matrix} \right\| . \qquad (65)$$

Thus for every unitary transformation there is a corresponding definite rotation of three-dimensional space, and all rotations may be obtained in this way. The product of two unitary transformations corresponds to the product of the corresponding rotations. We can say that expressions (59) define the homomorphism of the group of unitary transformations with unity determinant with the group of rotations of three-dimensional space.

We now consider what unitary transformations yield the identity transformation, i.e. the identity element in the rotation group. The third of expressions (59) gives us here

$$a\bar{b} = 0; \quad a\bar{a} - b\bar{b} = 1,$$

whence $|a| = 1$ and $b = 0$. Let $a = e^{i\delta}$. The first of (59) gives

$$\frac{1}{2}(e^{i2\delta} + e^{-i2\delta}) = 1.$$

It follows at once that $\delta = 0$ or π, i.e. $a = \pm 1$.

We thus have two unitary transformations with matrices

$$E = \left\| \begin{matrix} 1, & 0 \\ 0, & 1 \end{matrix} \right\|; \quad S = \left\| \begin{matrix} -1, & 0 \\ 0, & -1 \end{matrix} \right\| = -E,$$

to which the identity element of the rotation group corresponds.

Now suppose that two unitary transformations U and V give the same rotation. In this case, $V^{-1}U$ will give the identity transformation in the rotation group, i.e. $V^{-1}U = E$ or $(-E)$, i.e. $U = V$ or $U = -V$. We remark here that the $(-)$ sign in front of a matrix means that the signs of all the elements must be changed. The above discussion shows that unitary transformations (57) only lead to the same space rotation when they differ simply in sign. Conversely, if they only differ in sign, they give the same rotation, as shown

above and as also follows from (59). We can finally say that *the rotation group is homomorphic with the unitary transformation group with unity determinant, the same rotations being obtained when and only when the unitary matrices differ only in sign.*

The matrices E and $(-E)$ form a normal subgroup H of the group G of unitary transformations (57) with unity determinant. Any coset with respect to the normal subgroup H consists of two elements G_1 and $(-G_1)$, where G_1 is any element of the group. It follows at once from what was said above that *the rotation group is isomorphic to the complementary group to H.*

Expressions (59) contain two complex parameters a and b which must satisfy relationship (58). Each of the complex parameters contains two real parameters,

$$a = a_1 + ia_2;\ \ b = b_1 + ib_2,$$

and (58) is equivalent to the following:

$$a_1^2 + a_2^2 + b_1^2 + b_2^2 = 1.$$

Expressions (59) thus contain four real parameters which must satisfy a single relationship, i.e. (59) contain three independent real parameters, as must be the case for the rotation group. The parameters a and b are generally known as the *Cayley—Klein parameters.* It is easy to obtain their expression in terms of the Eulerian angles. For multiplication of the three unitary matrices (65) gives us, as we saw above, the unitary matrix which corresponds to the rotation with Eulerian angles $\{\alpha, \beta, \gamma\}$. On carrying out the multiplication, we obtain the following expressions for the corresponding parameters a and b:

$$a = e^{-i\frac{\alpha+\gamma}{2}}\cos\frac{\beta}{2};\ b = -ie^{i\frac{\gamma-\alpha}{2}}\sin\frac{\beta}{2}. \tag{66}$$

If 2π is added to α or γ, a and b change sign whilst the rotation remains in essence the same. This matter has already been mentioned above.

64. The general linear group and the Lorentz group. We have just established the close connection between the unitary group in two variables and the group of three-dimensional rotations. Similarly, a connection can be established between the general linear group in two variables with unity determinants, and the Lorentz group.

We introduce four variables

$$x_1, x_2, x_3, x_0,$$

and, on returning to expressions (51) for the stereographic projection, set in these

$$x = \frac{x_1}{x_0};\ y = \frac{x_2}{x_0};\ z = \frac{x_3}{x_0}. \tag{67}$$

This gives us the following expressions:

$$\frac{x_1}{x_0} = \frac{\bar{\xi}\eta + \xi\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}};\ \frac{x_2}{x_0} = \frac{1}{i}\frac{\bar{\xi}\eta - \xi\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}};\ \frac{x_3}{x_0} = \frac{\xi\bar{\xi} - \eta\bar{\eta}}{\xi\bar{\xi} + \eta\bar{\eta}}.$$

These define the x_k up to a common factor, and we can write

$$\left.\begin{aligned} x_0 &= \xi\bar{\xi} + \eta\bar{\eta}; \qquad x_1 = \bar{\xi}\eta + \xi\bar{\eta} \\ x_2 &= \frac{1}{i}(\bar{\xi}\eta - \xi\bar{\eta});\ x_3 = \xi\bar{\xi} - \eta\bar{\eta}. \end{aligned}\right\} \tag{68}$$

The previous variables satisfied relationship (51_1), and hence, by (67), the new variables given by (68) will satisfy for any complex ξ and η the expression

$$x_1^2 + x_2^2 + x_3^2 - x_0^2 = 0. \tag{69}$$

In the case of a unitary transformation from ξ and η, the expression $(\xi\bar{\xi} + \eta\bar{\eta})$ remained invariable, i.e. by (68), the variable x_0, here expressing time, remained invariable, and we thus obtained a rotation of three-dimensional space. We now abandon unitariness and take the general group of linear transformations

$$\xi' = a\xi + b\eta;\ \eta' = c\xi + d\eta. \tag{70}$$

We use a similar approach to that for unitary transformations. We form the expressions

$$\left.\begin{aligned} x_1 + ix_2 &= 2\bar{\xi}\eta;\ x_1 - ix_2 = 2\xi\bar{\eta}; \\ x_0 + x_3 &= 2\xi\bar{\xi};\ x_0 - x_3 = 2\eta\bar{\eta}. \end{aligned}\right\} \tag{71}$$

For the new variables ξ', η' we obtain new x_k':

$$\begin{aligned} x_1' + ix_2' &= 2\overline{\xi'}\eta';\ x_1' - ix_2' = 2\xi'\overline{\eta'} \\ x_0' + x_3' &= 2\xi'\overline{\xi'};\ x_0' - x_3' = 2\eta'\overline{\eta'}. \end{aligned}$$

On substituting expressions (70) and using (71), we get

$$\left.\begin{aligned} x_1' + ix_2' &= \bar{a}d\,(x_1 + ix_2) + \bar{b}c\,(x_1 - ix_2) + \\ &\quad + \bar{a}c\,(x_0 + x_3) + \bar{b}d\,(x_0 - x_3) \\ x_1' - ix_2' &= b\bar{c}\,(x_1 + ix_2) + a\bar{d}\,(x_1 - ix_2) + \\ &\quad + a\bar{c}\,(x_0 + x_3) + b\bar{d}\,(x_0 - x_3), \\ x_0' + x_3' &= \bar{a}b\,(x_1 + ix_2) + a\bar{b}\,(x_1 - ix_2) + \\ &\quad + a\bar{a}\,(x_0 + x_3) + b\bar{b}\,(x_0 - x_3), \\ x_0' - x_3' &= \bar{c}d\,(x_1 + ix_2) + c\bar{d}\,(x_1 - ix_2) + \\ &\quad + c\bar{c}\,(x_0 + x_3) + d\bar{d}\,(x_0 - x_3), \end{aligned}\right\} \tag{72}$$

whence linear expressions which we shall omit may be obtained with real coefficients for the x_k' in terms of the x_k. We merely observe that, if the last two of equations (72) are added, the coefficient of x_0 in the expression for x_0' turns out to be equal to $^1/_2\,(a\bar{a} + b\bar{b} + c\bar{c} + d\bar{d})$, i.e. the coefficient is positive.

The new variables satisfy a similar relationship to the old:

$$x_1'^2 + x_2'^2 + x_3'^2 - x_0'^2 = 0. \tag{73}$$

If we replace the x_k' on the left-hand side of this equation by their expressions in terms of the x_k, (69) must be obtained. But it is possible for the left-hand side of (73) to differ from the left-hand side of (69) by a constant factor, i.e. we have here

$$x_1'^2 + x_2'^2 + x_3'^2 - x_0'^2 = k\,(x_1^2 + x_2^2 + x_3^2 - x_0^2),$$

where k is a constant. On using the above expressions and taking into account the fact that

$$x_1'^2 + x_2'^2 + x_3'^2 - x_0'^2 = (x_1' + ix_2')\,(x_1' - ix_2') - (x_0' + x_3')\,(x_0' - x_3'),$$

it is easily shown that $k = (ad - bc)\,(\overline{ad} - \overline{bc}) = |\,ad - bc\,|^2$. If we want to have $k = 1$, i.e. the Lorentz transformation

$$x_1'^2 + x_2'^2 + x_3'^2 - x_0'^2 = x_1^2 + x_2^2 + x_3^2 - x_0^2, \tag{74}$$

we have to take linear transformations (70) with determinants of unit modulus, i.e. of the form $e^{i\varphi}$. On multiplying all the coefficients of (70) by $e^{-i\varphi/2}$ as before, on the one hand, we do not change the x_1', x_2', x_3' defined by (68) with ξ and η replaced by ξ' and η', since expressions

(68) contain the product of one of the quantities (ξ', η') with one of $(\bar{\xi}', \bar{\eta}')$, and on the other hand, we reduce the determinant to unity.

We therefore take transformations (70) as having unity determinant:

$$ad - bc = 1. \tag{75}$$

We can show as in the previous section that the linear transformation giving the x_k' in terms of the x_k has a $(+1)$ determinant. We recall furthermore that the coefficient of x_0 in the expression for x_0' is positive here, i.e. the transformation has a $(+1)$ determinant and does not change the direction for measuring time, or in other words, (72) yield positive Lorentz transformations.

To sum up, linear transformations under condition (75) give the positive Lorentz transformations which we defined in [54].

As in the previous section, we now pose the question of whether any given positive Lorentz transformation can be obtained from (72). We remark first of all that, as in the previous section, corresponding to the product of two linear transformations (70) we have the product of the corresponding Lorentz transformations, or more precisely: if A and B are two linear transformations (70) which lead in accordance with (72) to Lorentz transformations T_1 and T_2, corresponding to the linear transformation BA we have the Lorentz transformation $T_2 T_1$. As we saw in [54], every positive Lorentz transformation can be written in the form

$$T = VSU,$$

where U and V are simple rotations of three-dimensional space and S is a positive Lorentz transformation in two variables. In accordance with the results of the previous section, we can obtain any rotation with the aid of a unitary transformation of type (70) with unity determinant. It thus remains for us to show that we can get any positive Lorentz transformation S in two variables from (72), given a suitable choice of linear transformation (70). On comparing (74) with (21) of [54], it will be seen that we now take $c = 1$, so that expressions (17) of [54] for the positive Lorentz transformations in two variables become

$$\left.\begin{array}{c} x_3' = \dfrac{-vx_0 + x_3}{\sqrt{1-v^2}};\ x_0' = \dfrac{x_0 - vx_3}{\sqrt{1-v^2}} \\ x_1' = x_1;\ x_2' = x_2. \end{array}\right\} \tag{76}$$

We bring in the quantity

$$u = \frac{1}{\sqrt{1-v^2}} > 1$$

and take the particular form of transformation (70):

$$\xi' = l\xi;\ \eta' = \frac{1}{l}\eta,$$

where l is a real constant. The determinant of this is evidently unity. In the present case $a = l$, $d = 1/l$, and $b = c = 0$. On making these substitutions in (72), we in fact obtain (76) if l satisfies the conditions:

$$\frac{l^2}{2} + \frac{1}{2l^2} = u;\quad \frac{l^2}{2} - \frac{1}{2l^2} = -vu.$$

This at once gives us $l^2 = u \pm \sqrt{u^2 - 1}$. The second condition shows that we have to take the root less than unity for l^2 with $v > 0$, and the root greater than unity with $v < 0$, the second condition being thereby fulfilled. On extracting the root, we get two values of opposite sign for l. We can say finally that *the group of linear transformations* (70) *with determinant 1 is homomorphic with the group of positive Lorentz transformations, the homomorphism being in accordance with* (72). As in the previous section, this homomorphism is not an isomorphism, i.e. different transformations (70) can lead to the same Lorentz transformation. It follows at once from (72) that the identity transformation in the Lorentz group is obtained from the two linear transformations with matrices

$$E = \begin{Vmatrix} 1, 0 \\ 0, 1 \end{Vmatrix};\ S = \begin{Vmatrix} -1, & 0 \\ 0, & -1 \end{Vmatrix} = -E,$$

and it can be shown precisely as in the previous section that *any transformation of the Lorentz group can simply be obtained from two linear transformations* (70) *whose coefficients differ only in sign.*

As in [63], the elements E and $(-E)$ form a normal subgroup H of the group of linear transformations with determinant 1, and *the group of positive Lorentz transformations is isomorphic with the group complementary to H.*

Linear transformations (70) contain four complex coefficients, related by condition (75). Expressions (72) thus contain three arbitrary complex parameters, or in other words, six arbitrary real parameters.

65. Representation of a group by linear transformations. Let G be a group with elements G_a and suppose that there is a definite matrix

A_α corresponding to each G_α, all the A_α having the same order and non-zero determinants. Suppose further that the correspondence is such that the product $A_{\alpha_2} A_{\alpha_1}$ of matrices A_{α_2} and A_{α_1} corresponds to the product $G_{\alpha_2} G_{\alpha_1}$. We say in this case that *the matrices A_α or the corresponding linear transformations give a linear representation of the group G.* Let G_0 be the identity element of the group and A_0 the corresponding matrix. Since $G_0 G_\alpha = G_\alpha$, we must have $A_0 A_\alpha = A_\alpha$, whence, on multiplying on the right by A_α^{-1}, we have $A_0 = I$, i.e. the identity element must correspond to the unit matrix. Let G_{α_1} and G_{α_2} be inverse, and A_{α_1}, A_{α_2} the corresponding matrices. It follows from $G_{\alpha_2} \cdot G_{\alpha_1} = G_0$ that $A_{\alpha_2} A_{\alpha_1} = I$, i.e. inverse matrices correspond to inverse elements. An immediate consequence of the above is that *the matrices A_α (or the corresponding linear transformations) form a group A homomorphic to group G.* If distinct matrices correspond to distinct elements of G, A is isomorphic as well as homomorphic to G. It is said in this case to give a *one-to-one linear representation of group G.*

If this is not the case, the set of elements of G, to which the unit matrix in A corresponds, forms a normal subgroup of G, and group A is isomorphic to the group complementary to this normal divisor [57].

If the basic group G is itself a group of linear transformations, a possible linear representation is yielded by the group itself.

We notice one point in connection with the definition of linear representation. Suppose we know that to every element G_α there corresponds a definite matrix A_α, and that a product of matrices corresponds to a product of elements, but that we are unaware of whether the determinants of the A_α vanish or not. We show that, if one determinant $D(A_{\alpha_0})$ vanishes, all the $D(A_\alpha)$ vanish. Now the set of matrices $A_{\alpha_0} A_\alpha$ with variable α contains all the matrices corresponding to elements of the group [56]. But $D(A_{\alpha_0} A_\alpha) = D(A_{\alpha_0}) D(A_\alpha)$ and the product vanishes since the first factor vanishes by hypothesis. Hence, given a correspondence in which products correspond to products, we only need to verify that one of the determinants $D(A_\alpha)$ is non-zero; we only need to verify say that the unit matrix of A corresponds to the identity element of G.

Let X be a matrix of the same order as the A_α with a non-zero determinant. We have

$$(XA_{\alpha_2} X^{-1})(XA_{\alpha_1} X^{-1}) = XA_{\alpha_2} A_{\alpha_1} X^{-},$$

and consequently the matrices $XA_\alpha X^{-1}$ also give a linear representation of the group G. *Two such similar representations are generally described as equivalent.* Let the order of the A_α be n, and let $(x_1, \ldots, x_n)$ be vector components in n-dimensional space on which the transformations A_α are carried out, so that the group A becomes

$$\mathbf{x}' = A_\alpha \mathbf{x}. \tag{77}$$

As we know from [25], the equivalent linear representation

$$\mathbf{y}' = XA_\alpha X^{-1}\mathbf{y}, \tag{78}$$

means that new axes are taken in the space, the new components being given in terms of the old by

$$(y_1, \ldots, y_n) = X(x_1, \ldots, x_n) \tag{79}$$

With these new axes, linear transformations of space are now expressed by (78), i.e. the equivalent linear representations can be obtained by a simple change of the coordinate axes in accordance with (79). The variables $(x_1, \ldots, x_n)$ appearing in (77) are known as *the objects of the linear representation.* Passage to the equivalent linear representation is thus equivalent to replacing the objects of the linear representation by different objects with the aid of linear transformation (79) with non-zero determinant.

Let matrices A_α of order n give a linear representation of a group G, and let matrices B_α of order m give another linear representation of the same group. We form the quasi-diagonal matrix of order $(n+m)$:

$$[A_\alpha, B_\alpha] = \begin{Vmatrix} A_\alpha, & 0 \\ 0, & B_\alpha \end{Vmatrix}. \tag{80}$$

We have by the rule for multiplying quasi-diagonal matrices:

$$[A_{\alpha_2}, B_{\alpha_2}] \, [A_{\alpha_1}, B_{\alpha_1}] = [A_{\alpha_2}A_{\alpha_1}, B_{\alpha_2}B_{\alpha_1}].$$

Thus matrices (80) also give a linear representation of G. In general, given representations of G by means of matrices A_α, B_α, C_α, we can form a new representation by using the quasi-diagonal matrix

$$D_\alpha = [A_\alpha, B_\alpha, C_\alpha] = \begin{Vmatrix} A_\alpha, & 0, & 0 \\ 0, & B_\alpha, & 0 \\ 0, & 0, & C_\alpha \end{Vmatrix}, \tag{81}$$

We now observe that, if we pass to an equivalent representation by matrices $XD_\alpha X^{-1}$, the quasi-diagonal nature of the matrices is

in general destroyed, and we can no longer say at once from the form of this new representation that it is composed, up to an equivalent representation, of different representations with a smaller number of dimensions, in accordance with rule (81). If our linear representation D_a has the purely quasi-diagonal form (80), it breaks down into a number of linear representations A_a and B_a with a smaller number of dimensions, i.e. with matrices of a smaller order. The linear representation is said to be *reduced* in this case. If a linear representation E_a does not have a quasi-diagonal form but some equivalent representation $XE_a X^{-1}$ has such a form, E_a is said to be *reducible*. Finally, if neither the representation itself nor any representation equivalent to it has the quasi-diagonal form, i.e. neither is reduced, the representation is said to be *irreducible*.

We observe some conditions in which a representation can be said to be reduced. Let a linear representation consist of matrices A_a of order n which yield linear transformations in the variables $(x_1, \ldots, x_n)$. We suppose that all the A_a are unitary, and that the subspace R', formed by the first k fundamental vectors, is transformed into itself by the A_a, i.e. if $x_{k+1} = x_{k+2} = \ldots = x_n = 0$, then also $x'_{k+1} = x'_{k+2} = \ldots = x'_n = 0$. In other words, all the A_a have the form

$$\left\| \begin{matrix} A'_a, & N_a \\ 0, & A''_a \end{matrix} \right\|, \tag{82}$$

where A'_a is of order k, A''_a is of order $(n-k)$, and the bottom left corner, with $(n-k)$ rows and k columns, is filled with zeros. We consider the subspace R'' formed by the last $(n-k)$ fundamental vectors. It will consist of vectors orthogonal to all the vectors of the above R'. Since each A_a transforms R' into itself, and being unitary, preserves the vector orthogonality, each vector of R'' must become a vector likewise belonging to R'' as a result of the transformation A_a. In other words, if $x_1 = \ldots = x_k = 0$, then also $x'_1 = \ldots = x'_k = 0$. It follows at once from this that all the elements in the top right corner of (82), with k rows and $(n-k)$ columns, also vanish, i.e. the matrices of the linear representation in question are

$$\left\| \begin{matrix} A'_a, & 0 \\ 0, & A''_a \end{matrix} \right\| = [A'_a, \; A''_a],$$

and the representation is consequently reduced. Now let all the unitary transformations A_a leave completely invariant a subspace R_1 of k dimensions $(k < n)$, where n is the order of the A_a. We transform the

coordinate axes in such a way that R_1 is formed by the first k fundamental vectors, which is the same as passing to an equivalent linear representation and may be accomplished with the aid of a unitary transformation. By the above, the representation becomes reduced after this transformation. We thus have the following theorem:

THEOREM. *If a linear representation of a group consists of unitary matrices which leave some subspace unchanged, the representation is reducible.*

The reducibility or otherwise of a representation is closely bound up with the passage from matrices A_α to the similar matrices $XA_\alpha X^{-1}$. We shall notice some particular cases of passage to equivalent representations that may be obtained by a special choice of matrices X. We form a matrix X as follows: the first row has unity in the second place and zeros elsewhere, the second row has unity in the first place and zeros elsewhere, and from the third row onwards, we have unity on the principal diagonal and zeros elsewhere. Thus

$$X = \begin{Vmatrix} 0, & 1, & 0, & 0, & \ldots, & 0 \\ 1, & 0, & 0, & 0, & \ldots, & 0 \\ 0, & 0, & 1, & 0, & \ldots, & 0 \\ 0, & 0, & 0, & 1, & \ldots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0, & 0, & 0, & 0, & \ldots, & 1 \end{Vmatrix}.$$

We see on expanding directly, starting with the last row, that $D(X) = -1$. It may easily be verified by using the ordinary rules of matrix multiplication that, if Y is a given matrix, the similar matrix XYX^{-1} is obtained from Y by interchanging its first and second rows and its first and second columns. In the same way, every interchange of rows accompanied by a similar interchange of columns is equivalent to passage to a similar matrix with the aid of the transformation X, which is clearly independent of Y. Hence, *if we carry out the same interchange of rows and columns in all the matrices A_α yielding a linear representation of a group, this is equivalent to passing to an equivalent representation.*

If there exists a distribution of the integers $1, 2, \ldots, n$ into two classes such that, for each A_α, a zero stands at the intersection of a row numbered by an integer of one class with a column numbered by an integer of the other class, the representation is reducible. For its reduction is simply accomplished by interchanging rows and columns in such a way that those numbered from one class always stand

at the top and to the left, whilst those numbered by the other class stand below and to the right.

We conclude the present section by noticing the further case when the linear representation of a group G is of the first order, i.e. when all the A_a are first order matrices, or in other words, ordinary numbers. In this case, to each group element G_a there corresponds a transformation $x' = m_a x$, i.e. simply the number m_a, and the ordinary numerical product $m_2 m_1$ corresponds to the product $G_2 G_1$.

66. Basic theorems. Let G be a finite group consisting of m elements $G_1, \ldots, G_m$, and let $A_1, \ldots, A_m$ be nth order matrices that give a linear representation of the group. We write the objects of this representation as $\mathbf{x}(x_1, \ldots, x_n)$. We consider the expression

$$\varphi(x_1, \ldots, x_n) = \sum_{s=1}^{m} | A_s \mathbf{x} |^2. \tag{83}$$

This has the expanded form

$$\varphi = \sum_{s=1}^{m} \sum_{i=1}^{n} (a_{i1}^{(s)} x_1 + \ldots + a_{in}^{(s)} x_n)(\overline{a}_{i1}^{(s)} \overline{x}_1 + \ldots + \overline{a}_{in}^{(s)} \overline{x}_n), \tag{84}$$

where we have written $a_{ik}^{(s)}$ for the elements of matrix A_s. We can easily show that (84) is an Hermitian form, i.e. the coefficients of $\overline{x}_p x_q$ and $x_p \overline{x}_q$ are complex conjugates. Furthermore, it follows from (83) that this Hermitian form represents the sum of the squares of the lengths of certain vectors, i.e. the form is positive definite [40]. In other words, on carrying out the unitary transformation

$$\mathbf{y} = U\mathbf{x},$$

reducing our form to a sum of squares:

$$\varphi = \sum_{i=1}^{n} \lambda_j \overline{y}_j y_j,$$

all the coefficients λ_j will be positive. On carrying out the further transformation $z_j = \sqrt{\lambda_j} y_j$, we get an expression for the Hermitian form φ as a sum of pure squares:

$$\varphi = \overline{z}_1 z_1 + \ldots + \overline{z}_n z_n. \tag{85}$$

We subject the variables $(x_1, \ldots, x_n)$ to the transformation

$$\mathbf{x}' = A_k \mathbf{x}, \tag{86}$$

belonging to the linear representation of our group. It may easily be seen that, with this, the form φ remains unchanged.

In fact:

$$\varphi(x_1', \ldots, x_n') = \sum_{s=1}^{m}{}' \mid A_s A_k \mathbf{x} \mid^2 .$$

But, as we know from [56], the set of transformations (matrices)

$$A_1 A_k, A_2 A_k, \ldots, A_m A_k$$

is the same as the set

$$A_1, A_2, \ldots, A_m;$$

therefore, if we express transformation (86) in new variables ($z_1, \ldots, z_n$), related to the old by an expression of the form

$$(z_1, \ldots, z_n) = B_0(x_1, \ldots, x_n) ,$$

where B_0 is a matrix, we get instead of the group A_k the similar group $B_0 A_k B^{-1}$, and none of the transformations of this similar group will change (85), i.e. change the sum of the squares of the moduli; or in other words, all these transformations are unitary. We have thus shown that, for finite groups, every linear representation is equivalent to some unitary representation, i.e. a representation consisting of unitary transformations. This property is preserved, with certain supplementary conditions, for linear representations of parametrically dependent infinite groups, and in future, when a linear representation of a group is mentioned, we shall always understand it to be unitary. We now have the following theorem.

THEOREM I. *Every linear representation of a (finite) group has an equivalent unitary representation.*

We now find the necessary and sufficient condition for the reducibility of a linear representation. We first introduce a new term and call a diagonal matrix $[k, \ldots, k]$ with the same elements on the diagonal a *scalar matrix. Such a matrix may be written kI.* As we have seen above, it is equivalent to the number k as regards algebraic operations.

Suppose we are given a reducible linear representation of a group. The representation will be accomplished say by matrices of the form

$$D_a = X[A_a, B_a, C_a] X^{-1} ,$$

where the interior matrix is quasi-diagonal and X is a given matrix. We form the matrix

$$Y = X[kI, lI, mI] X^{-1} ,$$

where the quasi-diagonal middle term has the same structure as in the D_a. It may easily be seen that Y commutes with all the D_a. For

$$D_a Y = X[A_a k, B_a l, C_a m] X^{-1},$$

and similarly

$$YD_a = X[kA_a, lB_a, mC_a] X^{-1}.$$

But the order of the factors plays no part when a matrix is multiplied by a number. Furthermore, if the numbers k, l, m are different, as we shall assume, Y is not a scalar matrix. For it clearly has distinct characteristic roots k, l, m. We thus arrive at the following theorem.

THEOREM II. *If a linear representation is reducible, there exists a matrix differing from a scalar matrix that commutes with all the matrices appearing in the representation.*

We now show that the converse is also true, i.e.

THEOREM III. *If there exists a matrix Y which is not a scalar matrix and which commutes with all the matrices D_a of a linear representation, the representation is reducible.*

We have for any subscript a, by the conditions of the theorem:

$$D_a Y = YD_a. \tag{87}$$

Let Z be a matrix with a non-zero determinant such that all the matrices $ZD_a Z^{-1}$ are unitary: $ZD_a Z^{-1} = U_a$. We re-write the above equation as

$$Z^{-1} U_a ZY = YZ^{-1} U_a Z.$$

We obtain by multiplying on the left by Z and on the right by Z^{-1}:

$$U_a(ZYZ^{-1}) = (ZYZ^{-1}) U_a,$$

i.e. ZYZ^{-1} commutes with all the matrices of the unitary representation. This is clearly not a scalar matrix since if $ZYZ^{-1} = kI$, we have $Y = kI$. It is sufficient for us to prove the reducibility of the equivalent linear representation, U_a, and the proof of the theorem is thus reduced to the case when the representation is unitary. We simplify the writing by assuming that the linear representation consisting of matrices D_a is itself unitary.

Let λ_1 be a characteristic root of matrix Y. We know from [25] that the matrix $\lambda_1 = \lambda_1 I$ commutes with any matrix; consequently the matrix $Y - \lambda_1 I$ as well as Y satisfies condition (87), i.e. commutes with all the D_a. It is easily seen that at least one of the charac-

teristic roots of $Y_1 = Y - \lambda_1 I$ vanishes. For, the characteristic equation for Y_1 will be

$$D(Y_1 - \lambda I) = D[Y - (\lambda + \lambda_1) I] = 0,$$

i.e. it is found from the characteristic equation for Y by replacing λ by $(\lambda + \lambda_1)$, and since one of the characteristic roots of Y is λ_1, at least one of the characteristic roots of Y_1 is zero. It follows that the determinant of the matrix Y_1, equal to the product of the characteristic roots, also vanishes. We can thus assume in the proof of our theorem that all the D_a are unitary and that the determinant of the matrix Y appearing in (87) is zero.

We consider a set of vectors having the components

$$\left.\begin{array}{l} x_1 = y_{11} u_1 + y_{12} u_2 + \cdots + y_{1n} u_n \\ x_2 = y_{21} u_1 + y_{22} u_2 + \cdots + y_{2n} u_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n = y_{n1} u_1 + y_{n2} u_2 + \cdots + y_{nn} u_n , \end{array}\right\} \tag{88}$$

where the u_s take any values and the y_{ik} are elements of matrix Y. Since the determinant of Y vanishes, the rank of the array $\| y_{ik} \|$ is less than n. Let the rank be $r < n$. We know from [16] that in this case (88) define an r-dimensional subspace R'.

We consider the left-hand side of the equation

$$D_a Y \mathbf{u} = Y D_a \mathbf{u} , \tag{89}$$

The vector $Y\mathbf{u}$ has in fact the components (88), and $D_a Y \mathbf{u}$ is therefore the result of applying the transformation D_a to some arbitrarily chosen vector of the subspace R'. We have on the right-hand side of (89) the result of applying the transformation Y to the vector $D_a \mathbf{u}$, i.e. the components of the right-hand side are given by the same expressions (88) except that $u_1, \ldots, u_n$ are replaced by the components of $D_a \mathbf{u}$, i.e. the right-hand side of (89) represents a vector belonging to the sub-space R'. We see from this, on comparing the two sides, that the application of transformation D_a to any vector of R' yields a vector also belonging to R'. But we know from [65] that, if unitary transformations leave a subspace unchanged, they form a reducible representation. The theorem is thus proved.

Theorems II and III show that *the necessary and sufficient condition for the irreducibility of a linear representation is that there exists no matrix, not of the form kI, commuting with all the matrices appearing in the representation.*

It follows at once from Theorem I that there is no need to mention the unitariness of the representation in the theorem of [65], and it can be generally stated that, if all the matrices of a representation leave some subspace unchanged, the representation is reducible. The converse is obvious.

67. Abelian groups and representations of the first degree. A group G is described as *Abelian* if any two of its elements commute, i.e. $G_{a_2}G_{a_1}$ is equal to $G_{a_1}G_{a_2}$ for any subscripts [56]. Let A_{a_1}, A_{a_2} be the matrices corresponding to G_{a_1} G_{a_2} in a linear representation. We have A_{a_2} A_{a_1} corresponding to the product G_{a_2} G_{a_1}, and similarly A_{a_1} A_{a_2} to G_{a_1} G_{a_2}. But the products are the same and we must therefore have

$$A_{a2}\, A_{a1} = A_{a1}\, A_{a2},$$

i.e. *any two of the matrices of a linear representation of an Abelian group commute.*

Suppose the representation is unitary, i.e. all the matrices are unitary. We know that there now exists a unitary transformation U such that all the matrices $UA_a U^{-1}$ have a purely diagonal form [42]; thus there is an equivalent linear representation consisting of purely diagonal matrices

$$UA_a U^{-1} = [k_1^{(a)}, \ldots, k_n^{(a)}]\,.$$

It can be seen from this that the linear representation here decomposes into n first degree representations

$$B_a^{(s)} = k_s^{(a)} \qquad (s = 1, 2, \ldots, n)\,.$$

To sum up, *every unitary representation of an Abelian group is equivalent to a set of first degree representations, passage to an equivalent representation being likewise accomplished with the aid of a unitary transformation.*

We now consider some examples, both of Abelian group representations, and of first degree linear representations of non-Abelian groups.

Example **1.** We take as our first example the cyclic (Abelian) group of order m, consisting of the elements

$$S^0 = I, S, S^2, \ldots, S^{m-1} \qquad (S^m = I). \tag{90}$$

If the linear transformation $x' = \omega x$, or what amounts to the same thing, the number ω corresponds to the element S, we shall have the following numbers

corresponding to elements (90):

$$1, \omega, \omega^2, \ldots, \omega^{m-1}.$$

Since $S^m = I$, we must have $\omega^m = 1$, i.e.

$$\omega = e^{\frac{2k\pi i}{m}},$$

where k is a positive integer that can clearly be set equal to any one of 0, 1, 2, . . ., $m - 1$.

We consider in detail the case $m = 2$. Here we have

$$I, S \text{ and } S^2 = I,$$

i.e. $S = S^{-1}$. With $k = 0$, the identity transformation $x' = x$ or the number 1 corresponds to both the elements I and S; with $k = 1$, the transformation $x' = x$ corresponds to I, and $x' = -x$ to S, or in other words, the number 1 corresponds to I, and (-1) to S. An important case for physics is that when the group consists of the identity transformation of three-dimensional space and the symmetry transformation with respect to the origin:

$$x' = -x; \quad y' = -y; \quad z' = -z(S).$$

We clearly have $m = 2$. The two representations above may be termed the identity and alternating representation of symmetry with respect to the origin.

Example **2.** We take the group of rotations about the z axis. The matrices of the group have the form

$$Z_\varphi = \left\| \begin{matrix} \cos\varphi, & -\sin\varphi \\ \sin\varphi, & \cos\varphi \end{matrix} \right\| \tag{91}$$

and also satisfy, as we saw above, the obvious relationships

$$Z_{\varphi_2} Z_{\varphi_1} = Z_{\varphi_1} Z_{\varphi_2} = Z_{\varphi_1+\varphi_2}.$$

The function $e^{l\varphi}$ also satisfies these relationships. But it must be noticed that, if $\varphi = 2\pi$, the rotation is equivalent to the identity transformation, and we must therefore have $e^{2\pi l} = 1$, i.e. the number l must be of the form $l = ki$, where k is an integer. We thus have an infinite set of linear representations of the rotation group, with the numbers

$$e^{\varphi ki}$$

corresponding to matrices (91).

On assigning all possible values to the number k:

$$k = 0, \pm 1, \pm 2, \ldots,$$

we get the infinite set of linear representations of the rotation group.

Example 3. We now take the group consisting of the $n!$ permutations of n elements. We can associate the number $(+1)$ with each permutation, in which case we have what is known as the symmetric representation of the permutation group. Alternatively, we can associate the number $(+1)$ with even permutations, consisting of an even number of transpositions, and the number (-1) with odd permutations. This gives us what is known as the anti-symmetric representation of the permutation group. In this representation, the number $(+1)$ corresponds to each permutation of the alternating subgroup, and (-1) to the remaining permutations. It can be shown, though we shall not dwell on the proof, that the above two cases exhaust the possibilities as regards first degree linear representation of the permutation group. The group has other representations of higher degree than the first.

Example 4. We next take the group of all the real orthogonal transformations on a plane i.e. the group consisting of rotations of the plane about the origin combined with symmetrical reflection in the y axis. We saw above [52] that the matrices of this group are of the form

$$\{\varphi, d\} = \left\| \begin{matrix} d\cos\varphi, & -d\sin\varphi \\ \sin\varphi, & \cos\varphi \end{matrix} \right\|, \tag{92}$$

where $d = 1$ for a pure rotation and $d = -1$ for rotation combined with reflection. Apart from the obvious first degree linear representation in which each matrix (92) has the corresponding number $(+1)$, we can form a further first degree representation in which the number $(+1)$ corresponds to matrix (92) if $d = 1$, and the number (-1) corresponds to (92) if $d = -1$. This in fact gives us a linear representation, since the product of two matrices of form (92) corresponds to a pure rotation if d has the same sign in both matrices, and to a rotation with reflection if d has different signs.

68. Linear representations of the unitary group in two variables.

We consider the linear representations of the unitary group in two variables. As we know, this group has the form

$$\begin{aligned} x_1' &= ax_1 + bx_2 \\ x_2' &= -\bar{b}x_1 + \bar{a}x_2, \end{aligned} \tag{93}$$

where the complex numbers a and b must be subjected to the condition

$$a\bar{a} + b\bar{b} = 1.$$

We form the $(m+1)$ quantities:

$$\xi_0 = x_1^m;\ \xi_1 = x_1^{m-1}x_2;\ \ldots;\ \xi_m = x_2^m.$$

If we take $\xi'_k = x_1'^{m-k} x_2'^k$ and substitute for x'_1 and x'_2 from (93), each ξ'_k is clearly given linearly in terms of ξ_k, and hence we shall have for every transformation of group (93) a corresponding linear transformation from variables ξ_k to ξ'_k. It is obvious that products of transformations correspond to products of transformations, and we thus have a linear representation of group (93) of order $(m + 1)$. Though this representation may not be unitary, all we need do to get a unitary representation is to introduce an additional constant factor into each of variables (95), i.e. instead of (95), we define our variables by

$$\eta_k = \frac{x_1^{m-k} x_2^k}{\sqrt{(m-k)!k!}} \qquad (k = 0, 1, \ldots, m) \tag{96_1}$$

and similarly

$$\eta'_k = \frac{x_1'^{m-k} x_2'^k}{\sqrt{(m-k)!k!}} \qquad (k = 0, 1, \ldots, m) \tag{96_2}$$

where we reckon $0! = 1$ as usual.

We show that our representation is unitary with this definition of the variables, i.e.

$$\sum_{k=0}^{m} \eta'_k \overline{\eta}'_k = \sum_{k=0}^{m} \eta_k \overline{\eta}_k. \tag{97}$$

We have, in fact, on applying the binomial formula:

$$m! \cdot \sum_{k=1}^{m} \eta'_k \overline{\eta}'_k = m! \sum_{k=0}^{m} \frac{x_1'^{m-k}\, \overline{x_1'^{m-k}}\, x_2'^k\, \overline{x_2'^k}}{(m-k)!k!} = (x'_1 \overline{x'_1} + x'_2 \overline{x'_2})^m,$$

and similarly

$$m! \cdot \sum_{k=0}^{m} \eta_k \overline{\eta}_k = (x_1 \overline{x}_1 + x_2 \overline{x}_2)^m.$$

But since transformation (93) is unitary,

$$x'_1 \overline{x}'_1 + x'_2 \overline{x}'_2 = x_1 \overline{x}_1 + x_2 \overline{x}_2.$$

and relationship (97) therefore holds.

We now introduce explicit expressions for the coefficients of our unitary representation of group (93). We make a slight change in the notation for this purpose and put

$$\eta_l = \frac{x_1^{j+l} x_2^{j-l}}{\sqrt{(j+l)!\,(j-l)!}} \quad (l = -j,\ -j+1, \ldots, j-1,\ j). \tag{98}$$

In the previous notation $m = 2j$, and the number j will be an integer if m is even, and half an odd integer if m is odd. Let $m = 5$, expressions (98) give us the following six variables:

$$\eta_{-\frac{5}{2}} = \frac{x_2^5}{\sqrt{5!}}; \quad \eta_{-\frac{3}{2}} = \frac{x_1 x_2^4}{\sqrt{1!\,4!}}; \quad \eta_{-\frac{1}{2}} = \frac{x_1^2 x_2^3}{\sqrt{2!\,3!}};$$

$$\eta_{\frac{1}{2}} = \frac{x_1^3 x_2^2}{\sqrt{3!\,2'}}; \quad \eta_{\frac{3}{2}} = \frac{x_1^4 x_2}{\sqrt{4!\,1!}}; \quad \eta_{\frac{5}{2}} = \frac{x_1^5}{\sqrt{5!}}.$$

The variables here are enumerated, not by the first six integers, but by fractions that differ from each other by unity and run from $(-5/2)$ to $(+5/2)$. If $m = 4$, we have by (98) the five variables

$$\eta_{-2} = \frac{x_2^4}{\sqrt{4!}}; \quad \eta_{-1} = \frac{x_1 x_2^3}{\sqrt{1!\,3!}}; \quad \eta_0 = \frac{x_1^2 x_2^2}{\sqrt{2!2!}}; \quad \eta_1 = \frac{x_1^3 x_2}{\sqrt{3!\,1!}};$$

$$\eta_2 = \frac{x_1^4}{\sqrt{4!}}.$$

The variables here are enumerated by the integers from (-2) to $(+2)$. For every fixed $m = 2j$, we have precisely the same enumeration of rows and columns in the matrices which form the linear representation of order $(2j + 1)$ of group (93).

We next turn our attention to finding the elements of these matrices. We have

$$\eta_l' = \frac{x_1'^{\,j+l} x_2'^{\,j-l}}{\sqrt{(j+l)!\,(j-l)!}} = \frac{(ax_1 + bx_2)^{j+l}(-\bar{b}x_1 + \bar{a}x_2)^{j-l}}{\sqrt{(j+l)!\,(j-l)!}}$$

and we want to write the right-hand side as a linear combination of the quantities η_l. Application of the binomial formula gives us

$$\eta' = \sum_{k=0}^{j+l} \sum_{k'=0}^{j-l} (-1)^{j-l-k'} \frac{\sqrt{(j+l)!\,(j-l)!}}{k!k'!\,(j+l-k)!\,(j-l-k')!} \times$$

$$\times\ \bar{a}^{k'} a^{j+l-k}\, \bar{b}^{j-l-k'}\, b^k\, x_1^{2j-k-k'} x_2^{k+k'}.$$

If we reckon $p! = \infty$ when p is a negative integer, we can take the summation with respect to k and k' in the above expression from $(-\infty)$ to $(+\infty)$, since the superfluous terms will contain an infinite factor

in the denominator and therefore vanish. We replace k' by a new variable of summation $s = j - k - k'$, so that the summation is over integral or half integral values from $(-\infty)$ to $(+\infty)$, depending on whether j is an integer or half an integer. We thus get

$$\eta_l' = \sum_k \sum_s (-1)^{k+s-l} \frac{\sqrt{(j+l)!\,(j-l)!}}{k!(j-k-s)!\,(j+l-k)!\,(k+s-l)!} \times$$
$$\times\, \bar{a}^{j-k-s} a^{j+l-k} \bar{b}^{k+s-l} b^k x_1^{j+s} x_2^{j-s}.$$

But we have by (98):

$$x_1^{j+s} x_2^{j-s} = \sqrt{(j+s)!\,(j-s)!}\,\eta_s,$$

and we finally get the required linear relationship in the form

$$\eta_l' = \sum_k \sum_s (-1)^{k+s-l} \frac{\sqrt{(j+l)!\,(j-l)!\,(j+s)!\,(j-s)!}}{k!\,(j-k-s)!\,(j+l-k)!\,(k+s-l)!} \times$$
$$\times\, \bar{a}^{j-k-s} a^{j+l-k} \bar{b}^{k+s-l} b^k \eta_s.$$

Thus, having assigned a fixed j, the elements of the matrix of the linear transformation of order $(2j+1)$, corresponding to unitary transformation (93) with matrix

$$\left\| \begin{matrix} a, & b \\ -\bar{b}, & \bar{a} \end{matrix} \right\|,$$

will be

$$D_j\begin{Bmatrix} a, b \\ -\bar{b}, \bar{a} \end{Bmatrix}_{ls} = (-1)^{s-l} \sum_k (-1)^k \times$$
$$\times \frac{\sqrt{(j+l)!\,(j-l)!\,(j+s)!\,(j-s)!}}{k!\,(j-k-s)!\,(j+l-k)!\,(k+s-l)!} \bar{a}^{j-k-s} a^{j+l-k} \bar{b}^{k+s-l} b^k. \tag{99}$$

The subscripts l and s here run over the following values:

$$l \text{ and } s = -j, -j+1, \ldots, j-1, j,$$

where the further reminder must be given that, if j is half an integer, we have an enumeration of rows and columns of the matrix likewise in half integers. On taking $p! = \infty$ if p is a negative integer, we get the following limits of summation with respect to k in (99):

$$k \geqslant 0; \quad k \geqslant l - s; \quad k \leqslant j - s; \quad k \leqslant j + l. \tag{100}$$

We notice that (99) may be simplified by passing to a similar representation. Let A be a matrix with elements a_{pq} and S the diagonal matrix $[\delta_1, \ldots, \delta_n]$.

It is easily seen by using the ordinary multiplication rule that the matrix SAS^{-1} has the elements

$$\{SAS^{-1}\}_{pq} = \delta_p a_{pq} \delta_q^{-1}.$$

If we now apply this rule to matrices

$$D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a} \end{Bmatrix}$$

and take $\delta_k = (-1)^k$, the factor $(-1)^{s-l}$ goes out in (99); in fact, we shall assume below that this factor is absent.

We next turn to the proof that *the linear representation of unitary group* (93) *defined by matrices with elements* (99) *is irreducible.* We prove two preliminary lemmas.

LEMMA I. *If a diagonal matrix, of which no two of the diagonal elements are the same, commutes with a matrix A, A is also diagonal.*

We have by hypothesis:

$$A[\delta_1, \ldots, \delta_n] = [\delta_1, \ldots, \delta_n]\, A,$$

where no two of the δ_k are the same. Let a_{pq} be the elements of A. Using the multiplication rule, we get from the above:

$$a_{pq}\,\delta_q = \delta_p\, a_{pq} \text{ or } a_{pq}(\delta_q - \delta_p) = 0,$$

and consequently $a_{pq} = 0$ if $p \neq q$, i.e. A is in fact a diagonal matrix.

LEMMA II. *If a diagonal matrix $[\delta_1, \ldots, \delta_n]$ commutes with a matrix A in which at least one column contains no zeros, we have $\delta_1 = \ldots = \delta_n$.*

On interchanging rows and columns, i.e. passing to similar matrices, we can bring the column with no zeros into the first place. With this, the diagonal matrix still remains diagonal, and the matrices commute as before. We can thus suppose, writing a_{pq} for the elements of A, that

$$a_{i1} \neq 0 \qquad (i = 1, 2, \ldots, n),$$

and moreover, by hypothesis, as above:

$$a_{i1}(\delta_1 - \delta_i) = 0 \qquad (i = 1, 2, \ldots, n),$$

whence we have $\delta_1 = \ldots = \delta_n$, and the lemma is proved.

We now prove the irreducibility of the linear representation defined by matrices (99). Let Y be a matrix of order $(2j+1)$ which commutes with all the matrices

$$D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a} \end{Bmatrix},$$

obtained with differing a and b satisfying condition (94). To prove irreducibility, we want to show that Y must be a scalar matrix. We first take the case when $b = 0$ and $a = e^{i\alpha}$. These complex numbers clearly satisfy (94).

Using (99), we first of all find that now

$$D_j\begin{Bmatrix} e^{i\alpha}, & 0 \\ 0, & e^{-i\alpha} \end{Bmatrix}_{ls} = 0 \qquad l \neq s,$$

whilst the diagonal elements are

$$D_j\begin{Bmatrix} e^{i\alpha}, & 0 \\ 0, & e^{-i\alpha} \end{Bmatrix}_{ll} = e^{i2l\alpha} \quad (l = -j, \ -j+1, \dots, j-1, j),$$

and the matrix has the form

$$D_j\begin{Bmatrix} e^{i\alpha}, & 0 \\ 0, & e^{-i\alpha} \end{Bmatrix} = \begin{Vmatrix} e^{-i2j\alpha}, & 0, & 0, & \dots, 0 \\ 0, & e^{-i2(j-1)\alpha} & 0, & \dots, 0 \\ 0, & 0, & 0^{-i2(j-2)\alpha} & \dots, 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0, & 0, & 0, & \dots, e^{i2j\alpha} \end{Vmatrix}, \qquad (101)$$

i.e. given a suitable choice of a, we have a diagonal matrix with different elements on the principal diagonal. We can say, by using the first lemma, that the matrix Y that has to commute with matrices (101) must also be diagonal, i.e.

$$Y = [\delta_1, \dots, \delta_n]. \qquad (102)$$

We now take the case when both numbers a and b differ from zero, and consider the first column of the matrix $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a} \end{Bmatrix}$. Its elements are given by (99) on setting $s = -j$. Inequalities (100) now give us

$$k \geqslant 0; \ k \geqslant l+j; \ k \leqslant 2j; \ k \leqslant j+l \ (l = -j, \ -j+1, \dots, j-1, j),$$

whence it is clear that the entire sum appearing in (99) reduces to a single term, which is obtained with $k = j + l$ and evidently differs from zero. Thus the first column of the matrix $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a} \end{Bmatrix}$ does not in fact

contain zero. But since the diagonal matrix (102) must commute with this matrix, all the δ_k must be the same by Lemma II, i.e. Y is a scalar matrix. The matrices $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a}\end{Bmatrix}$ therefore in fact yield an irreducible linear representation of unitary group (93). On assigning to j the series of values

$$j = 0, \frac{1}{2}, \; 1, \; \frac{3}{2}, 2, \ldots$$

we get an infinite set of these linear representations. With $j = 0$ we get the trivial identity representation, when the number unity corresponds to every element of group (93). We now consider, with $j > 0$, to what transformation of group (93) there corresponds the identity transformation of the representation group $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a}\end{Bmatrix}$, which is defined by the equations $\eta_l' = \eta_l$ or, what amounts to the same thing, by the equations

$$(ax_1 + bx_2)^{j+l}(-\bar{b}x_1 + \bar{a}x_2)^{j-l} = x_1^{j+l}x_2^{j-l}$$
$$(l = -j\;\; -j, +1, \ldots, j-1, \; j).$$

Setting $j = l$, we get

$$(ax_1 + bx_2)^{2j} = x_1^{2j},$$

whence it follows that $b = 0$, and the previous equations may be written as

$$a^{j+l}\,\bar{a}^{j-l}x_1^{j+l}x_2^{j-l} = x_1^{j+l}x_2^{j-l} \qquad (l = -j, \; -j+1, \ldots, j-1, j),$$

whence $a^{j+l}\,\bar{a}^{j-l} = 1$. But $|a| = 1$ for $b = 0$, and the last equation may be rewritten

$$a^{2l} = 1 \qquad (l = -j, \; -j+1, \ldots, \; j-1, \; j).$$

If j is half an odd integer, we can put $l = 1/2$ which gives $a = 1$. If j is an integer, the equations $a^{2l} = 1$ reduce simply to $a^2 = 1$, whence $a = \pm 1$.

Hence, if j is half an odd integer, the identity transformation in the group $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a}\end{Bmatrix}$ merely corresponds to the identity transformation in group (93), i.e. in this case $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a}\end{Bmatrix}$ is a one-to-one representation of group (93). Whereas if j is an integer, to the identity transformation in the $D_j\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a}\end{Bmatrix}$ group there correspond two transformations

in group (93) with matrices

$$E = \begin{Vmatrix} 1, & 0 \\ 0, & 1 \end{Vmatrix}; \quad S = \begin{Vmatrix} -1, & 0 \\ 0, & 1 \end{Vmatrix} = -E.$$

These transformations form a second order cyclic group, and the $D_j \begin{Bmatrix} a, b \\ -\bar{b}, \bar{a} \end{Bmatrix}$ are a one-to-one representation of the complementary group [58]. We can say alternatively that, to every transformation in the $D_j \begin{Bmatrix} a, \bar{b} \\ -\bar{b}, \bar{a} \end{Bmatrix}$ representation with integral j there correspond two transformations of group (93), for which the numbers a and b differ only in sign.

69. Linear representations of the rotation group. The above results are of particular importance since the unitary group (93) is closely bound up with the group of rotations of three-dimensional space; in fact, we can use these results to obtain an irreducible linear representation of the rotation group.

We have a definite rotation corresponding to every unitary transformation (93), whilst a simultaneous change of sign of a and b gives a unitary transformation to which the same rotation corresponds. The parameters a and b are related to the Eulerian angles of the corresponding rotation by the expressions [63]:

$$a = e^{-\frac{1}{2} i(\gamma + \alpha)} \cos \frac{1}{2} \beta; \qquad b = -ie^{\frac{1}{2} i(\gamma - \alpha)} \sin \frac{1}{2} \beta. \tag{103}$$

We first take the case when j is an integer. Expressions (99) show us here that a simultaneous change of sign of a and b does not alter the terms on the right-hand side, since the sum of the exponents of a, $\bar{a}$, b and $\bar{b}$ is equal to the even number $2j$. Thus the same matrix in the linear representation corresponds here to the two unitary transformations that yield the same rotation. In other words, to each rotation with Eulerian angles $\{\alpha, \beta, \gamma\}$ there corresponds, with integral j, a definite matrix in the linear representation D_j. Instead of $D_j \begin{Bmatrix} a, b \\ -\bar{b}, \bar{a} \end{Bmatrix}$, we shall now write for this matrix

$$D_j\{\alpha, \beta, \gamma\}. \tag{104}$$

If j is half an odd integer, simultaneous change of sign of a and b leads to a change of sign of all expressions (99), i.e. we have here,

corresponding to the unitary transformations that lead to the same motion, different matrices in which all the elements are of different sign. Two matrices of different sign also correspond to each rotation, i.e. we now have to put signs in front of the D_j in (104). To sum up, matrices (104) give us a linear representation of the rotation group when j is an integer. When j is half an odd integer, we do not strictly speaking obtain a linear representation, but have what is termed a two-valued linear representation.

To obtain expressions for the elements of matrices (104), all we need do is substitute from (103) for a and b in (99). We obtain, on first neglecting the factor $(-1)^{s-l}$:

$$D_j\{\alpha, \beta, \gamma\}_{ls} = i^{s-l} \sum_k (-1)^k \frac{\sqrt{(j+l)!\,(j-l)!\,(j+s)!\,(j-s)!}}{k!(j-k-s)!\,(j+l-k)!\,(k+s-l)!} \times$$
$$\times\, e^{-il\alpha - is\gamma} \cos^{2j+l-2k-s}\frac{1}{2}\beta \sin^{2k+s-l}\frac{1}{2}\beta\,. \tag{105}$$

If we take advantage of passage to the equivalent representation with the aid of the matrix

$$X = \left\| \begin{matrix} 0, & 0, & \ldots, & 0, & 1 \\ 0, & 0, & \ldots, & 1, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0, & 1, & \ldots, & 0, & 0 \\ 1, & 0, & \ldots, & 0, & 0 \end{matrix} \right\|,$$

it becomes a question of interchanging rows and columns in the reverse order, i.e. of replacing the l and s by $(-l)$ and $(-s)$. We can thus write, instead of (105), the new expressions

$$D_j'\{\alpha, \beta, \gamma\}_{ls} = i^{l-s} \sum_k (-1)^k \frac{\sqrt{(j+l)!\,(j-l)!\,(j+s)!\,(j-s)!}}{k!(j-k+s)!\,(j-l-k)!\,(k-s+l)!} \times$$
$$\times\, e^{il\alpha - is\gamma} \cos^{2j-l-2k+s}\frac{1}{2}\beta \sin^{2k-s+l}\frac{1}{2}\beta\,, \tag{106}$$

the factor i^{l-s} being neglected by using the same arguments as in [68].

We note some elementary particular cases. With $j = 0$, we have the first degree linear representation

$$\eta' = \eta\,.$$

This is the trivial identity representation. With $j = 1/2$, we have $2j + 1 = 2$, and the quantities $\eta_{-\frac{1}{2}}$ and $\eta_{\frac{1}{2}}$ are simply equal to x_2 and x_1, i.e. unitary group (93) is here its own special linear representation (apart from possible interchange of rows and columns).

We obtain for the rotation group a two-valued representation of the second degree, defined by the matrices

$$D'_{\frac{1}{2}}\{a, \beta, \gamma\} = \left\| \begin{matrix} e^{-\frac{1}{2}i(\gamma+a)}\cos\frac{1}{2}\beta, & ie^{\frac{1}{2}i(\gamma-a)}\sin\frac{1}{2}\beta \\ ie^{-\frac{1}{2}i(\gamma-a)}\sin\frac{1}{2}\beta, & e^{\frac{1}{2}i(\gamma-a)}\cos\frac{1}{2}\beta \end{matrix} \right\|.$$

With $j = 1$, we have a linear representation of the third degree:

$$D'_1\{a, \beta, \gamma\} = \left\| \begin{matrix} e^{-i(\gamma+a)}\frac{1+\cos\beta}{2}, & -e^{-ia}\frac{\sin\beta}{\sqrt{2}}, & e^{i(\gamma-a)}\frac{1-\cos\beta}{2} \\ e^{-i\gamma}\frac{\sin\beta}{\sqrt{2}}, & \cos\beta, & -e^{i\gamma}\frac{\sin\beta}{\sqrt{2}} \\ e^{-i(\gamma-a)}\frac{1-\cos\beta}{2}, & e^{ia}\frac{\sin\beta}{\sqrt{2}}, & e^{i(\gamma+a)}\frac{1+\cos\beta}{2} \end{matrix} \right\|.$$

The linear representations $D'_j\ \{a, \beta, \gamma\}$ with integral j give one-to-one representations of the rotation group. This follows at once from the fact that two matrices of group (93) correspond to each $D_j\ \{a, \beta, \gamma\}$, where the matrices differ only in the signs of a and b and, as mentioned above, correspond with the same rotation. If j is half an odd integer, to each rotation there correspond two matrices of the $D_j\ \{a, \beta, \gamma\}$ representation, differing only in sign. In particular, matrices $\pm E$ of the $D_j\ \{a, \beta, \gamma\}$ correspond to the identity transformation of the rotation group, where E is the unit matrix of order $(2j + 1)$. *If we confine ourselves to transformations of $D_j\ \{a, \beta, \gamma\}$ sufficiently close to the identity transformation, the $D_j\ \{a, \beta, \gamma\}$ are single-valued representations of the rotation group.* In this case, it is sufficient to confine ourselves to values of a, β, γ, near enough to zero in the general expressions (106). But if we add 2π to a or γ, all the matrices of $D_j\ \{a, \beta, \gamma\}$ change sign in view of the fact that s and l are half odd numbers, and we get a second representation of essentially the same rotation. We show later that the *representations mentioned are all the isomorphic irreducible representations of the rotation group.*

Since the $D_j\ \{a, \beta, \gamma\}$ are all irreducible representations of the rotation group, the matrix $D'_1\{a, \beta, \gamma\}$ must be similar to $D\{a, \beta, \gamma\}$, corresponding to the rotation of space with Eulerian angles $\{a, \beta, \gamma\}$. We saw in [63] that $D = Z_a X_\beta Z_\gamma$, and on carrying out the matrix

multiplication, we get

$$D = \begin{vmatrix} \cos\alpha\cos\gamma - \sin\alpha\cos\beta\sin\gamma, & -\cos\alpha\sin\gamma - \sin\alpha\cos\beta\cos\gamma, & \sin\alpha\sin\beta \\ \sin\alpha\cos\gamma + \cos\alpha\cos\beta\sin\gamma, & -\sin\alpha\sin\gamma + \cos\alpha\cos\beta\cos\gamma, & -\cos\alpha\cos\beta \\ \sin\beta\sin\gamma, & \sin\beta\cos\gamma, & \cos\beta \end{vmatrix},$$

and it may easily be verified that

$$AD_1'\{\alpha, \beta, \gamma\} A^{-1} = D\{\alpha, \beta, \gamma\},$$

where

$$A = \begin{Vmatrix} 1, & 0, & 1 \\ i, & 0, & -i \\ 0, & \sqrt{2}i & 0 \end{Vmatrix}.$$

70. The theorem on the simplicity of the rotation group. We show next that the rotation ~~group is~~ simple, i.e. it has no normal subgroups [58]. If there were such a subgroup, it would follow from what was said in [63] that there exists a corresponding normal subgroup of group G of transformations (57) with unity determinant, differing from the normal subgroup H consisting of E and $(-E)$. We therefore want to show that group G has no normal subgroup other than H, i.e. that if a normal subgroup H_1 of G contains a matrix A different from E and $(-E)$, H_1 coincides with G. We observe first of all that, if H_1 contains a matrix B, it follows by the definition of normal divisor that H_1 contains all the matrices $U^{-1}BU$, where U is any matrix of group G. We can thus obtain by suitable choice of U any matrix of G having the same characteristic roots as B. Hence to show that H_1 is the same as G, we only need to show that H_1 contains a matrix with any permissible characteristic roots. These roots must have the form $e^{i\omega}$ and $e^{-i\omega}$, where ω is a real number, since the matrix is unitary and its determinant is equal to unity.

By what has been said, we can take $U^{-1}AU$ instead of the matrix A and it can therefore be assumed that A is diagonal.

Let it be given then that H_1 contains $A = [e^{i\varphi}, e^{-i\varphi}]$, where φ is real, and $e^{i\varphi} \neq \pm 1$. With this, $A^{-1} = [e^{-i\varphi}, e^{i\varphi}]$. We take the arbitrary matrix of group G:

$$U = \begin{Vmatrix} x, & y \\ -\bar{y}, & \bar{x} \end{Vmatrix} \qquad (\bar{x}x + y\bar{y} = 1).$$

Now,

$$U^{-1} = \begin{Vmatrix} \bar{x}, & -y \\ \bar{y}, & x \end{Vmatrix}.$$

Since the subgroup H_1 contains A and is a normal divisor, it must also contain the matrix

$$Y = A(UA^{-1}U^{-1}).$$

On carrying out the matrix multiplication and taking into account that $x\bar{x} + y\bar{y} = 1$, we get the following expression for the trace s of Y:

$$s = 2 - 4y\bar{y}\sin^2\varphi = 2 - 4\varrho^2\sin^2\varphi,$$

where $\varrho = |y|$ can take any value from the interval $0 \leqslant \varrho \leqslant 1$, and $\sin\varphi \neq 0$. The characteristic roots of Y, $(e^{i\alpha}, e^{-i\alpha})$, are roots of the equation

$$\lambda^2 - s\lambda + 1 = 0, \quad \text{i.e.} \quad \lambda^2 + (4\varrho^2\sin^2\varphi - 2)\lambda + 1 = 0.$$

As ϱ varies from $\varrho = 0$ to $\varrho = 1$, α runs from $\alpha = 0$ to $\alpha = 2\varphi$. We introduce the following notation:

$$U_\beta = [e^{i\beta}, e^{-i\beta}].$$

It follows from the above remarks that H_1 contains all the matrices H_α with $0 \leqslant \alpha \leqslant 2\varphi$. It is now easy to show that H_1 contains any matrix U_β $(\beta > 0)$. For on choosing a positive integral n such that $0 < \beta/n < 2\varphi$, H_1 will contain $U_{\beta/n}$, and will therefore also contain

$$U^n_{\frac{\beta}{n}} = U_\beta.$$

Hence H_1 contains matrices with any characteristic roots and by what we have said above, must coincide with G. We have thus shown that *the rotation group is simple*.

It follows at once from this that the rotation group cannot have homomorphic (as distinct from isomorphic) representations. For if there were such a representation, to the identity transformation in the representation group there would have to correspond in the rotation group transformations forming a normal subgroup, whereas no normal subgroup exists by the above.

71. Laplace's equation and linear representations of the rotation group. We shall next indicate the connection between linear representations of groups and differential equations. This connection lies at the basis of the application of linear representations to problems of modern physics. We shall start with the elementary case of Laplace's equation [II, 92]: whilst giving us nothing new, this will throw some light on the subject as a whole. We first establish some general facts

that play an important part in problems of linear representations of groups; these generalizations are indeed already familiar to us in certain particular cases from the above examples.

Let G, whose linear representations we wish to form, be a group of linear transformations of order n:

$$x'_k = g_{k1}^{(a)} x_1 + \dots + g_{kn}^{(a)} x_n \quad (k = 1, 2, \dots, n), \tag{107}$$

where the superscript a, characterizing an element of G, runs over a finite or infinite set of values. Further, let m functions exist:

$$\varphi_s(x_1, \dots, x_n) \qquad (s = 1, 2, \dots, m), \tag{108}$$

such that they also undergo linear transformation on substituting for the independent variables in accordance with (107):

$$\varphi_s(x'_1, \dots, x'_n) = a_{s1}^{(a)} \varphi_1(x_1, \dots, x_n) + \dots + a_{sm}^{(a)} \varphi_m(x_1, \dots, x_n) \tag{109}$$
$$(s = 1, 2, \dots, m).$$

We have here a matrix A_a with elements $a_{ik}^{(a)}$ corresponding to transformation (107) of group G. We consider two transformations of the group:

$$(x'_1, \dots, x'_n) = G_{a1}(x_1, \dots, x_n); \; (x''_1, \dots, x''_n) = G_{a2}(x'_1, \dots, x'_n);$$
$$G_{a3} = G_{a2} G_{a1}.$$

The corresponding transformations of functions (108) are

$$\varphi_s(x'_1, \dots, x'_n) = a_{s1}^{(a_1)} \varphi_1(x_1, \dots, x_n) + \dots + a_{sm}^{(a_1)} \varphi_m(x_1, \dots, x_n) \tag{110$_1$}$$

and

$$\varphi_s(x''_1, \dots, x''_n) = a_{s1}^{(a_2)} \varphi_1(x'_1, \dots, x'_n) + \dots + a_{sm}^{(a_2)} \varphi_m(x'_1, \dots, x'_n). \tag{110$_2$}$$

On substituting for the $\varphi_s(x'_1, \dots, x'_n)$ from (110_1) in (110_2), we get a direct relationship for $\varphi_s(x''_1, \dots, x''_n)$ in terms of $\varphi_s(x_1, \dots, x_n)$ yielding a matrix A_{a_3}. We thus obtain

$$\{A_{a_3}\}_{ik} = \sum_{s=1}^{m} a_{is}^{(a_2)} a_{sk}^{(a_1)}, \quad \text{i.e.} \quad A_{a_3} = A_{a_2} A_{a_1},$$

and expressions (109) evidently define a linear representation of order m of group G. We have assumed in the above arguments that the functions φ_s are linearly independent. In this case linear transformations (109) are uniquely defined and $D(A_a) \neq 0$, since otherwise the $\varphi_s(x'_1, \dots, x'_n)$ would be connected by a linear relationship.

In the particular case of constructing linear representations of a unitary group, the role of functions φ_s was played by functions (96_1).

Let G be the group of rotations of three-dimensional space, so that $n = 3$, and let the φ_s be orthogonal and normalized in a sphere K with centre at the origin, i.e.

$$\iiint_K \varphi_p(x_1, x_2, x_3)\, \overline{\varphi_q(x_1, x_2, x_3)}\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3 = \delta_{pq}. \tag{111}$$

We show that linear representation (109) of the rotation group is now unitary. The sphere K is displaced into itself as a result of a rotation G_α, and the determinant of G_α is known to be equal to unity. Condition (111) thus gives us

$$\iiint_K \varphi_p(x_1', x_2', x_3')\, \overline{\varphi_q(x_1', x_2', x_3')}\, \mathrm{d}x_1'\, \mathrm{d}x_2'\, \mathrm{d}x_3' = \delta_{pq},$$

or by (109):

$$\iiint_K \Big[\sum_{i=1}^m a_{pi}^{(\alpha)} \varphi_i(x_1, x_2, x_3) \cdot \sum_{j=1}^m \overline{a_{qj}^{(\alpha)} \varphi_j(x_1, x_2, x_3)}\Big]\, \mathrm{d}x_1'\, \mathrm{d}x_2'\, \mathrm{d}x_3' = \delta_{pq}.$$

By the rule for change of variables in a triple integral, on passing to (x_1, x_2, x_3) we have simply to substitute $\mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3$ for $\mathrm{d}x_1'\, \mathrm{d}x_2'\, \mathrm{d}x_3'$ then integrate over the same sphere K. We get by (111):

$$\sum_{i=1}^m a_{pi}^{(\alpha)} \overline{a_{qi}^{(\alpha)}} = \delta_{pq} \qquad (p, q = 1, 2, \ldots, m),$$

where, as usual, $\delta_{pq} = 0$ for $p \neq q$ and $\delta_{pp} = 1$, i.e. each of the matrices A is here orthogonal by rows, whilst the transposed matrices will be orthogonal by columns and consequently by rows also [28]; it follows that a basic matrix has orthogonality both of rows and columns, or in other words, the A_α are in fact unitary matrices for any α.

We now consider the Laplace equation in two variables

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \tag{112}$$

or, in vector notation,

$$\operatorname{div} \operatorname{grad} U = 0. \tag{113}$$

We take the homogeneous polynomial in x and y of degree l:

$$\varphi_l(x, y) = a_0 x^l + a_1 x^{l-1} y + \ldots + a_k x^{l-k} y^k + \ldots + a_l y^l. \tag{114}$$

We show that there exist two linearly independent polynomials of type (114) that are solutions of equation (112), and that every solution

of (112) that is a homogeneous poynomial of degree l must be a linear combination of the above two polynomials with constant coefficients. In fact, the coefficients of polynomial (114) are given by

$$a_k = \frac{1}{(l-k)!\,k!}\,\frac{\partial^l \varphi_l(x, y)}{\partial x^{l-k}\,\partial y^k}.$$

But since this polynomial must satisfy equation (112), we can replace double differentiation with respect to y by double differentiation with respect to x whilst at the same time changing the sign, inasmuch as (112) can be written as

$$\frac{\partial^2 U}{\partial y^2} = -\frac{\partial^2 U}{\partial x^2}.$$

We thus obtain for the coefficients a_k expressions of the form

$$a_k = \pm\frac{1}{(l-k)!\,k!}\,\frac{\partial^l \varphi_l}{\partial x^l} \quad \text{or} \quad a_k = \pm\frac{1}{(l-k)!\,k!}\,\frac{\partial^l \varphi_l}{\partial x^{l-1}\,\partial y},$$

i.e. all the coefficients of polynomial (114) are expressible in terms of coefficients a_0 and a_1. This argument shows us that there exist no more than two linearly independent homogeneous polynomials satisfying equation (112). We now show that two such distinct polynomials in fact exist. For this, we consider the homogeneous polynomial

$$\omega_l(x, y) = (x + iy)^l.$$

After removing the brackets and separating real and imaginary parts, we get

$$\omega_l(x, y) = \varphi_l(x, y) + i\psi_l(x, y),$$

where $\varphi_l(x, y)$ and $\psi_l(x, y)$ are real linearly independent homogeneous polynomials of degree l. We get by differentiating $\omega_l(x, y)$:

$$\frac{\partial^2 \omega_l(x, y)}{\partial x^2} = l(l-1)(x+iy)^{l-2};$$

$$\frac{\partial^2 \omega_l(x, y)}{\partial y^2} = -l(l-1)(x+iy)^{l-2},$$

i.e. $\omega_l(x, y)$ satisfies equation (112). The same can therefore be said of the real and imaginary parts of this function, i.e. of polynomials $\varphi_l(x, y)$ and $\psi_l(x, y)$, and these give us the two required solutions of (112). We introduce polar coordinates

$$x = r\cos\varphi; \quad y = r\sin\varphi,$$

whence

$$\omega_l(x, y) = r^l e^{il\varphi}.$$

Polynomials φ_l and ψ_l now take the very simple forms

$$\varphi_l(x, y) = r^l \cos l\varphi; \quad \psi_l(x, y) = r^l \sin l\varphi.$$

We rotate the xy plane about the origin by the angle:

$$x' = x\cos\vartheta - y\sin\vartheta, \quad y' = x\sin\vartheta + y\cos\vartheta. \tag{115}$$

It is easily seen that equation (112) now remains invariant, or more precisely, it looks exactly the same in the new variables:

$$\frac{\partial^2 U}{\partial x'^2} + \frac{\partial^2 U}{\partial y'^2} = 0. \tag{116}$$

This can be verified directly by using (115) and the rule for differentiating a function of a function. Or it follows directly from the fact that the left-hand side of equation (113) has a definite value independently of the axes chosen, so that it has the same form for any Cartesian axes. The polynomials $\varphi_l(x', y')$, $\psi_l(x', y')$ must satisfy equation (116), and consequently also (112), and must therefore be linearly expressible in terms of $\varphi_l(x, y)$ and $\psi_l(x, y)$. This in fact gives us a linear representation of the group of rotations on a plane.

We now take two different polynomials that are linear combinations of the above:

$$\varphi_l'(x, y) = \varphi_l(x, y) - i\psi_l(x, y); \quad \psi_l'(x, y) = \varphi_l(x, y) + i\psi_l(x, y)$$

or

$$\varphi_l'(x, y) = (x - iy)^l = r^l e^{-il\varphi}; \quad \psi_l'(x, y) = (x + iy)^l = r^l e^{il\varphi}.$$

These polynomials yield the following transformation:

$$\varphi_l'(x', y') = r^l e^{-il(\varphi+\vartheta)} = e^{-il\vartheta}\varphi_l'(x, y),$$
$$\psi_l'(x', y') = r^l e^{il(\varphi+\vartheta)} = e^{il\vartheta}\psi_l'(x, y),$$

i.e. the matrix

$$\begin{Vmatrix} e^{-il\vartheta}, & 0 \\ 0, & e^{il\vartheta} \end{Vmatrix}$$

corresponds in the linear representation to transformation (115), where the angle ϑ can have any value. The form of the matrix implies at once that the linear representation has a reduced form. It gives two linear representations of the first order, defined by the numbers $e^{-il\vartheta}$ and $e^{il\vartheta}$. The integer l can take any value throughout the above discussion. We have obtained in this way the same linear representations of the rotation group for a plane as were obtained earlier in [69].

We now turn to the Laplace equation in three variables:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = 0 \tag{117}$$

or

$$\operatorname{div} \operatorname{grad} U = 0. \tag{118}$$

We consider homogeneous polynomials of degree l, now in three variables:

$$\varphi_l(x, y, z) = a_0 z^l + X_1(x, y) z^{l-1} + X_2(x, y) z^{l-2} + \ldots + \\ + X_{l-1}(x, y) z + X_l(x, y), \tag{119}$$

where the $X_k(x, y)$ are homogeneous polynomials of degree k in their arguments. Each $X_k(x, y)$ contains $(k + 1)$ arbitrary coefficients, so that in general the polynomial $\varphi_l(x, y, z)$ of degree l in three variables will contain the following number of arbitrary coefficients:

$$1 + 2 + 3 + \ldots + (l + 1) = \frac{(l + 1)(l + 2)}{2}.$$

On substituting (119) in equation (117), we get a homogeneous polynomial of degree $(l - 2)$ on the left, and on equating its coefficients to zero, we get $l(l - 1)/2$ homogeneous equations in the $(l + 1)(l + 2)/2$ unknown coefficients of $\varphi_l(x, y, z)$. We have:

$$\frac{(l + 1)(l + 2)}{2} - \frac{(l - 1) l}{2} = 2l + 1,$$

so that at least $(2l + 1)$ coefficients in $\varphi_l(x, y, z)$ remain arbitrary, i.e. there will exist at least $(2l + 1)$ linearly independent homogeneous polynomials of degree l satisfying equation (117). By using the same method as for two variables, it can be shown that there are not more than $(2l + 1)$ of these polynomials, i.e. there are precisely $(2l + 1)$. We write these polynomials as

$$\varphi_s^{(l)}(x, y, z) \qquad (s = 1, 2, \ldots, 2l + 1).$$

If

$$(x', y', z') = U \cdot (x, y, z)$$

is a rotation of three-dimensional space about the origin, equation (117) will meantime remain invariant, and polynomials $\varphi_s^{(l)}(x, y, z)$ give a linear representation of order $(2l + 1)$ of the group of rotations of three-dimensional space.

We give later a detailed treatment of these harmonic polynomials and introduce explicit expressions for them. We shall see that they can always be chosen so as to be orthogonal and normalized in any sphere

with centre at the origin. The linear representation of the rotation group that they then afford is unitary. This representation can be shown to be in fact equivalent to the representation $D_l\{\alpha, \beta, \gamma\}$ which we constructed in [69]. We shall return to this problem later.

72. Direct matrix products. Suppose we have two matrices

$$A = \begin{Vmatrix} a_{11}, a_{12}, \ldots, a_{1n} \\ a_{21}, a_{22}, \ldots, a_{2n} \\ \cdots\cdots\cdots\cdots \\ a_{n1}, a_{n2}, \ldots, a_{nn} \end{Vmatrix} \text{ and } B = \begin{Vmatrix} b_{11}, b_{12}, \ldots, b_{1m} \\ b_{21}, b_{22}, \ldots, b_{2m} \\ \cdots\cdots\cdots\cdots \\ b_{m1}, b_{m2}, \ldots, b_{mm} \end{Vmatrix}, \qquad (120)$$

the first being of order n and the second of order m. We form a new matrix C, whose elements $c_{ij,kl}$ are obtained by multiplying each element of A by each element of B:

$$\{C\}_{ij;kl} = c_{ij;kl} = a_{ik} b_{jl}. \qquad (121)$$

Here the set of two integers (i, j) plays the role of first subscript, and the set of integers (k, l) that of second subscript, where

$$i \text{ and } k = 1,\ 2,\ \ldots,\ n;$$

$$j \text{ and } l = 1,\ 2,\ \ldots,\ m.$$

In other words, we have a special method of reference to rows and columns, in which they are indicated by a set of two integers, the first taking values from 1 to n and the second from 1 to m. We can naturally enumerate the rows and columns in the ordinary way by simple integers which go from 1 to nm, with one such definite integer corresponding to each pair of (i, j) or (k, l), the integers being the same if the pairs are the same. Various different methods can be used for the enumeration by single integers. Passing from one method to another amounts to a simultaneous interchange of rows and columns, i.e. to passage to a similar matrix which will later have no significance.

The matrix C is called *the direct product of matrices* A and B, and is generally written as

$$C = A \times B. \qquad (122)$$

The order of the factors is of no significance in this new type of product.

Suppose, for instance, that both matrices (120) are of the second order. Their direct product is now a matrix of the fourth order which

we can write say in the form

$$C = \begin{Vmatrix} a_{11}b_{11}, & a_{11}b_{12}, & a_{12}b_{11}, & a_{12}b_{12} \\ a_{11}b_{21}, & a_{11}b_{22}, & a_{12}b_{21}, & a_{12}b_{22} \\ a_{21}b_{11}, & a_{21}b_{12}, & a_{22}b_{11}, & a_{22}b_{12} \\ a_{21}b_{21}, & a_{21}b_{22}, & a_{22}b_{21}, & b_{22}b_{22} \end{Vmatrix} = \begin{Vmatrix} c_{11;11}, & c_{11;12}, & c_{11;21}, & c_{11;22} \\ c_{12;11}, & c_{12;12}, & c_{12;21}, & c_{12;22} \\ c_{21;11}, & c_{21;12}, & c_{21;21}, & c_{21,22} \\ c_{22;11}, & c_{22;12}, & c_{22;21}, & c_{22;22} \end{Vmatrix}$$

or in an alternative form, given a simultaneous interchange of rows and columns.

Let A and B be diagonal matrices:

$$A = [\gamma_1, \ldots, \gamma_n]; \quad B = [\delta_1, \ldots, \delta_m].$$

In this case $a_{ik} = 0$ and $b_{lj} = 0$ for $i \neq k$ and $j \neq l$, and consequently, by (121), $c_{ij;kl}$ only differs from zero when the pair (i, j) is the same as (k, l), i.e. the matrix C will also be diagonal. The principal diagonal contains all the possible products of the γ_k with the δ_l. If all the γ_k and δ_l are unity, C is also a unit matrix. We thus have the following theorem.

THEOREM I. *The direct product of two diagonal matrices is a diagonal matrix, and the direct product of two unit matrices is a unit matrix.*

We also prove the following theorem.

THEOREM II. *If $A^{(1)}$ and $A^{(2)}$ are two matrices of the same order n and $B^{(1)}$, $B^{(2)}$ are matrices of the same order m, the following formula is valid:*

$$\left(A^{(2)} \times B^{(2)}\right)\left(A^{(1)} \times B^{(1)}\right) = A^{(2)} A^{(1)} \times B^{(2)} B^{(1)}. \tag{123}$$

It should be noticed that when we write two matrices of the same order after each other with no sign, this always means the ordinary product of the matrices. Denoting the elements by the corresponding small letters with two subscripts, we have by definition of direct product:

$$\left\{A^{(t)} \times B^{(t)}\right\}_{ij;kl} = a_{ik}^{(t)} b_{jl}^{(t)} \qquad (t = 1,\ 2),$$

and on using the ordinary matrix multiplication rule, we get the following expression for an element of the left-hand side of (123):

$$d_{ij;kl} = \sum_{p=1}^{n} \sum_{q=1}^{m} a_{ip}^{(2)} b_{jq}^{(2)} a_{pk}^{(1)} b_{ql}^{(1)}. \tag{124}$$

We show that the same expression is obtained for the corresponding element of the right-hand side. We have by the definition of ordinary product:

$$\left\{A^{(2)} A^{(1)}\right\}_{ik} = \sum_{p=1}^{n} a_{ip}^{(2)} a_{pk}^{(1)}; \quad \left\{B^{(2)} B^{(1)}\right\}_{jl} = \sum_{q=1}^{m} b_{jq}^{(2)} b_{ql}^{(1)}.$$

and by definition of direct product:

$$d_{ij;kl} = \sum_{p=1}^{n} a_{ip}^{(2)} a_{pk}^{(1)} \cdot \sum_{q=1}^{m} b_{jq}^{(2)} b_{ql}^{(1)}$$

which is the same as (124). We now turn to the proof of a final theorem regarding direct products.

THEOREM III. *If A and B are unitary matrices, their direct product $C = A \times B$ is also unitary.*

We have by hypothesis:

$$\sum_{s=1}^{n} a_{sp} \overline{a}_{sq} = \delta_{pq}; \quad \sum_{s=1}^{m} b_{sp} \overline{b}_{sq} = \delta_{pq}. \tag{125}$$

We verify that the columns of C are orthogonal and normalized, and write

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij;\,p_1 q_1} \overline{c}_{ij;\,p_2 q_2} = \delta_{p_1 q_1;\, p_2 q_2}.$$

i.e. by (121):

$$\delta_{p_1 q_1;\, p_2 q_2} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ip_1} \overline{a}_{ip_2} b_{jq_1} \overline{b}_{jq_2} = \sum_{i=1}^{n} a_{ip_1} \overline{a}_{ip_2} \sum_{j=1}^{m} b_{jq_1} \overline{b}_{jq_2}. \tag{126}$$

If the pairs of numbers (p_1, q_1) and (p_2, q_2) are different, at least one of the factors on the right-hand side of (126) is zero, whilst if the pairs are the same, both factors are unity by (125). Hence $\delta_{p_1 q_1 p_2 q_2}$ is zero if the pairs are different, and unity if the pairs are the same, which proves the theorem.

We can clearly form the direct product of three matrices by forming the direct product of the direct product of the first two with the third:

$$A^{(1)} \times A^{(2)} \times A^{(3)}.$$

On retaining the previous notation, we have the following for an element of the new matrix:

$$c_{ikl;\,i'k'l'} = a_{ii'}^{(1)} a_{kk'}^{(2)} a_{ll'}^{(3)}.$$

The direct product of any finite number of matrices may be formed in a similar manner, the order of the matrix representing the direct product being the product of the orders of the matrix factors. The order of the factors is of no significance.

73. The composition of two linear representations of a group. Suppose we have two linear representations of a given group G with elements G_a:

$$x_i' = a_{i1}^{(a)} x_1 + \dots + a_{in}^{(a)} x_n \qquad (i = 1, 2, \dots, n) \tag{127}$$

and

$$y_k' = b_{k1}^{(a)} y_1 + \dots + b_{km}^{(a)} y_m \qquad (k = 1, 2, \dots, m) \tag{128}$$

where the superscript a runs over a finite or infinite set of values. We write $A^{(a)}$ and $B^{(a)}$ for the matrices of transformations (127) and (128) and form their direct prouct:

$$C^{(a)} = A^{(a)} \times B^{(a)}. \tag{129}$$

The matrices $C^{(a)}$ also give a linear representation of the group G. For to any element G_a of G there corresponds a matrix $C^{(a)}$; and the product $G_{a_2} G_{a_1} = G_{a_3}$ has the corresponding matrix $C^{(a_2)} C^{(a_1)}$ which is given, by (123), by

$$C^{(a_2)} C^{(a_1)} = \left(A^{(a_2)} \times B^{(a_2)}\right)\left(A^{(a_1)} \times B^{(a_1)}\right) = \left(A^{(a_2)} A^{(a_1)}\right) \times \left(B^{(a_2)} B^{(a_1)}\right).$$

But since matrices $A^{(a)}$ and $B^{(a)}$ give linear representations of the group, we have

$$A^{(a_2)} A^{(a_1)} = A^{(a_3)} \quad \text{and} \quad B^{(a_2)} B^{(a_1)} = B^{(a_3)},$$

and consequently:

$$C^{(a_2)} C^{(a_1)} = A^{(a_3)} \times B^{(a_3)},$$

i.e. by (129):

$$C^{(a_2)} C^{(a_1)} = C^{(a_3)}.$$

Thus to a product of elements G_a corresponds the product of corresponding matrices $C^{(a)}$, and these matrices give a new linear representation of G. We notice that we now have the direct product of unit matrices $A^{(a)}$ and $B^{(a)}$, i.e. a unit matrix $C^{(a)}$, corresponding to the identity element of group G.

We form the nm products $x_i y_k$ and subject each of the factors to transformations (127) and (128). We have

$$x_i' y_k' = (a_{i1}^{(a)} x_1 + \dots + a_{in}^{(a)} x_n) \cdot (b_{k1}^{(a)} y_1 + \dots + b_{km}^{(a)} y_m),$$

or on removing the brackets:

$$x_i' y_k' = \sum_{p=1}^{n} \sum_{q=1}^{m} c_{ik;\,pq}^{(a)} x_p y_q, \quad \text{where} \quad c_{ik;\,pq}^{(a)} = a_{ip}^{(a)} b_{kq}^{(a)},$$

i.e. *if x_i and y_k are objects in linear representations defined by matrices $A^{(a)}$ and $B^{(a)}$, then $x_i y_k$ are objects in the linear representation defined by*

matrices $C^{(a)}$. If the $A^{(a)}$ and $B^{(a)}$ give irreducible linear representations, the representation given by the $C^{(a)}$ is not necessarily irreducible. We treat later the case when G is the group of rotations of three-dimensional space, and the $A^{(a)}$ and $B^{(a)}$ are the different irreducible linear representations of the group that we constructed in [69]. We show that in this case the product

$$D_{j1}\{\alpha, \beta, \gamma\} \times D_{j2}\{\alpha, \beta, \gamma\}$$

is reducible, and we find the irreducible representations that compose it.

We take as an example the Schrödinger equation for two electrons in the field of a positive nucleus. The equation is of the form

$$\left[-\frac{h^2}{8\pi^2 m}\sum_{s=1}^{2}\left(\frac{\partial^2}{\partial x_s^2}+\frac{\partial^2}{\partial y_s^2}+\frac{\partial^2}{\partial z_s^2}\right)+V\right]\psi = E\psi, \tag{130}$$

where

$$V=\sum_{s=1}^{2}-\frac{e^2 e_0}{\sqrt{x_s^2+y_s^2+z_s^2}}+\frac{1}{2}\frac{e^2}{\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}} \tag{131}$$

the constants having their usual significance. The second term in the expression for V is due to the interaction of the electrons. If we neglect this interaction as a first approximation, the equation becomes

$$(H_1+H_2)\,\psi = E\psi, \tag{132}$$

where

$$H_s = -\frac{h^2}{8\pi^2 m}\left(\frac{\partial^2}{\partial x_s^2}+\frac{\partial^2}{\partial y_s^2}+\frac{\partial^2}{\partial z_s^2}\right)-\frac{e^2 e_0}{\sqrt{x_s^2+y_s^2+z_s^2}} \quad (s = 1, 2).$$

Suppose that the separate equations:

$$H_1\psi = E\psi; \quad H_2\psi = E\psi \tag{133}$$

have eigenvalues E_1 and E_2 and corresponding eigenfunctions

$$\psi_1(x_2, y_1, z_1) \quad \text{and} \quad \psi_2(x_2, y_2, z_2),$$

i.e.

$$H_1\psi_1 = E_1\psi_1 \quad H_2\psi_2 = E_2\psi_2. \tag{134}$$

If we substitute in (132):

$$\psi = \psi_1(x_1, y_1, z_1)\cdot\psi_2,(z_2, y_2, z_2),$$

we clearly get by (134):

$$(H_1+H_2)\,\psi = \psi_2 H_1\psi_1 + \psi_1 H_2\psi_2 = (E_1+E_2)\,\psi_1\psi_2 = (E_1+E_2)\,\psi,$$

i.e. equation (132) will have the eigenfunction $\psi_1\psi_2$, to which (E_1+E_2) corresponds. The left-hand sides of equations (133) contain Laplace's operator and the distance of a point from the origin, and they are consequently unchanged on carrying out a rotation of space about the origin. It may happen that more than one eigenfunction ψ_1 corresponds to the characteristic root $E = E_1$ in

the first of equations (133). In this case, all the eigenfunctions in question, representing solutions of the first of equations (133), yield a linear representation of the rotation group, just as did the homogeneous harmonic polynomials of [69]. Let the representation be $D_{j1}\{\alpha, \beta, \gamma\}$. In precisely the same way, the solutions of the second of equations (133) for a given eigenvalue $E = E_2$ give us a representation $D_{j2}\{\alpha, \beta, \gamma\}$ of the rotation group. According to the above, the product $\psi_1 \psi_2$ gives us a representation of the rotation group equal to the direct product $D_{j1} \times D_{j2}$, and to recognize the physical characteristics of the corresponding eigenvalue $(E_1 + E_2)$ of equation (132) it becomes essential for us to distinguish the component irreducible representations. This circumstance has a fundamental role in excitation theory.

74. The direct product of groups and its linear representations. The concept of the direct product of matrices plays a part in another problem to which we shall now turn our attention. Let G and H be two groups, with elements G_α and H_β, where the α and β run independently over in general different sets of values. We define a new group F with elements defined by pairs of elements of G and H:

$$F_{\alpha\beta} = (G_\alpha, H_\beta),$$

the first element being from G and the second from H. The identity (unit) element of the new group is defined as the $F_{\alpha\beta}$ when the G_α and H_β are the identity elements of G and H, and inverse elements of F are defined in a similar manner. We naturally define the multiplication rule for F by

$$F_{\alpha_2\beta_2} F_{\alpha_1\beta_1} = (G_{\alpha_2} G_{\alpha_1}, H_{\beta_2} H_{\beta_1}).$$

As is easily seen, the set of $F_{\alpha\beta}$ in fact forms a group. We call the group F the direct product of groups G and H. Suppose we have a linear representation of G formed by matrices $A^{(\alpha)}$ and a linear representation of H formed by matrices $B^{(\beta)}$. It can be shown by using (123), as in the previous section, that *the direct products*

$$C^{(\alpha, \beta)} = A^{(\alpha)} \times B^{(\beta)}$$

give a linear representation of group F. Moreover, if $A^{(\alpha)}$ and $B^{(\beta)}$ are unitary representations, the representation $C^{(\alpha, \beta)}$ of F is also unitary [72].

We now show that, *if the representations $A^{(\alpha)}$ and $B^{(\beta)}$ are irreducible, the representation $C^{(\alpha, \beta)}$ of group F is also irreducible*. Let the $A^{(\alpha)}$ be of order n, and the $B^{(\beta)}$ of order m. The matrices $C^{(\alpha, \beta)}$ will be of order nm. Let a matrix X exist of order nm which commutes with all the $C^{(\alpha, \beta)}$. Matrix elements will be denoted by the corresponding small

letters. We thus have, for any subscripts i, j, p, q, and for any α and β:

$$\sum_{l=1}^{m}\sum_{k=1}^{n} x_{ij;\,kl}\, a_{kp}^{(\alpha)} b_{lq}^{(\beta)} = \sum_{l=1}^{m}\sum_{k=1}^{n} a_{ik}^{(\alpha)} b_{jl}^{(\beta)} x_{kl;\,pq}, \tag{135}$$

where

$$a_{kp}^{(\alpha)} b_{lq}^{(\beta)} = c_{kl;\,pq}^{(\alpha,\,\beta)} \quad \text{and} \quad a_{ik}^{(\alpha)} b_{jl}^{(\beta)} = c_{ij;\,kl}^{(\alpha,\,\beta)}.$$

If we take $G^{(\alpha)}$ as the identity element of G, $A^{(\alpha)}$ will be a unit matrix, i.e. $a_{kp}^{(\alpha)} = 0$ for $k \neq p$ and $a_{pp}^{(\alpha)} = 1$, and (135) gives us:

$$\sum_{l=1}^{m} x_{ij;\,pl}\, b_{lq}^{(\beta)} = \sum_{l=1}^{m} b_{jl}^{(\beta)} x_{il,\,pq}, \tag{136}$$

and similarly, taking $H^{(\beta)}$ as the identity element of H, we get

$$\sum_{k=1}^{m} x_{ij;\,kp}\, a_{kp}^{(\alpha)} = \sum_{k=1}^{n} a_{ik}^{(\alpha)} x_{kj;\,pq}. \tag{137}$$

If we take nm elements $x_{ij;\,kl}$ and fix the subscripts i and k, we get the m^2 elements

$$x_{ij;\,kl} \qquad (j,\ l = 1, 2, \ldots, m)$$

which give a matrix of order m. We write $X^{(i,\,k)}$ for this matrix. Similarly, on fixing j and l in $x_{ij;\,kl}$, we get a matrix $X_2^{(j,\,l)}$ of order n, By (136), all the $X_1^{(i,\,k)}$ will commute with all the matrices $B^{(\beta)}$, forming an irreducible representation of group H, and consequently all the $X_1^{(i,\,j)}$ are scalar matrices, i.e. the elements $x_{ij;\,kl}$ for fixed i and k have the same value if $j = l$, and are zero if $j \neq l$. We can write this as follows:

$$x_{ij;\,kl} = x_{i1;\,k1}\,\delta_{jl}. \tag{138$_1$}$$

Similarly, we have by considering the matrices $X_2^{(j,\,l)}$:

$$x_{ij;\,kl} = x_{1j;\,1l}\,\delta_{ik}, \tag{138$_2$}$$

where, as usual,

$$\delta_{pq} = 0 \quad \text{for} \quad p \neq q \quad \text{and} \quad \delta_{pp} = 1.$$

It follows from equations (138$_1$) and (138$_2$) that $x_{ij;\,kl}$ differs form zero only when $i = k$ and $j = l$, in which case all the $x_{ij;\,kl}$ are numerically the same, i.e. the X commuting with all the $C^{(\alpha,\,\beta)}$ is necessarily a scalar matrix. It now follows immediately that the linear representation of group F defined by the direct product $A^{(\alpha)} \times B^{(\beta)}$ is irreducible. It can be shown that *all the irreducible representations of group F are obtained in this way.*

Let G and H be groups of linear transformations with the same number of variables, and let any pair of matrices G_α and H_β commute:

$$G_\alpha H_\beta = H_\beta G_\alpha. \tag{139}$$

We have assumed in the above arguments that an element of group F is defined by a pair of elements (G_α, H_β), and we have laid down a definite rule, written above, for multiplication within group F. In the present case we can take the elements of F as simply the matrix products (139) which are independent of the order of the factors. This new F is isomorphic with the previous group F. If G_{α_0} and H_{β_0} are unit matrices, the product $G_{\alpha_0} H_{\beta_0} = H_{\beta_0} G_{\alpha_0}$ is also a unit matrix. The matrix $G_\alpha^{-1} H_\beta^{-1} = H_\beta^{-1} G_\alpha^{-1}$ is clearly the inverse of $G_\alpha H_\beta$, and we have the following multiplication rule by (139):

$$G_{\alpha_2} H_{\beta_2} \cdot G_{\alpha_1} H_{\beta_1} = (G_{\alpha_2} G_{\alpha_1})(H_{\beta_2} H_{\beta_1}),$$

i.e. all the previous properties are satisfied here in the formation of F, so that product (139) can be taken as the variable element of group F. We take the particular case when G is the group of rotations of three-dimensional space and H is the second order group consisting of the identity transformation I and symmetry S with respect to the origin [57]. Condition (139) is satisfied here. If G_α is any rotation of space, clearly $G_\alpha S = SG_\alpha$. In this case F is the group of all real orthogonal transformations of three-dimensional space. We had two first degree linear representations in [67] for H. One was the identity representation consisting of the number $(+1)$, and the other was the anti-symmetric representation, in which $(+1)$ corresponds with the matrix I and (-1) with matrix S. If we now take a linear representation $D_j\{\alpha, \beta, \gamma\}$ of the rotation group, we can take the direct product of a matrix of this representation with both these representations of the group of symmetry with respect to the origin. We obtain in the one case a linear representation of the total group of orthogonal transformations in which the same matrix $D_j\{\alpha, \beta, \gamma\}$ corresponds to every rotation with Eulerian angles $\{\alpha, \beta, \gamma\}$, whether taken in the pure form or in association with symmetry with respect to the origin. We write $D_j^+\{\alpha, \beta, \gamma\}$ for this representation of the group of orthogonal transformations. In the second case, the matrix $D_j\{\alpha, \beta, \gamma\}$ corresponds to a pure rotation and $-D_j\{\alpha, \beta, \gamma\}$ to a rotation in association with symmetrical reflection. We write $D_j^-\{\alpha, \beta, \gamma\}$ for this latter representation of the orthogonal transformation group.

We shall discuss one further example of the direct product of two groups. Let (x_1, y_1, z_1) and (x_2, y_2, z_2) be two points and G the group of rotations of three-dimensional space. Our variables now undergo the linear transformations:

$$\begin{aligned} x'_k &= g_{11} x_k + g_{12} y_k + g_{13} z_k, \\ y'_k &= g_{21} x_k + g_{22} y_k + g_{23} z_k, \qquad (k = 1, 2) \\ z'_k &= g_{31} x_k + g_{32} y_k + g_{33} z_k, \end{aligned} \tag{140}$$

where the array of g_{ik} is the matrix of a certain rotation. We suppose further that H is the group consisting of the identity transformation and the transformation corresponding to interchange of the subscripts 1 and 2 in our points. This latter transformation will have the form

$$\begin{pmatrix} 1, & 2 \\ 2, & 1 \end{pmatrix} \qquad (S). \tag{141}$$

We obviously have $S^2 = I$, and the group H will therefore consist of the two transformations I and S. Given a rotation G_a, clearly $G_a S = S G_a$, since it is a matter of indifference whether the renumbering of the points comes before or after the rotation. We obtain here the same linear representations for the total group F as above. If we took n instead of two points, the group H, consisting of interchanges of the point subscripts, would have for its elements linear transformations in n variables, and H would be isomorphic with the group of permutations of n elements. In this last case, the operations of rotation and of point subscript permutation similarly commute with each other, and the direct products of matrices of the linear representation of the rotation group with matrices of the linear representation of the permutation group give us a linear representation of the total group F.

75. Decomposition of the composition $D_j \times D_{j'}$ of linear representations of the rotation group. We now return to our discussion of [73] of the Schrödinger equation for two electrons, where we saw that, neglecting electron interaction, the eigenfunctions of the Schrödinger equation give us a linear representation of the rotation group which is obtained by the composition of two linear representations of this group. The results of the previous section show that it is important for us to be able to decompose such a linear representation into its irreducible parts. This is the problem that concerns us in the present article, and it may be stated mathematically as follows. Suppose we have two irreducible

representations $D_j\{\alpha, \beta, \gamma\}$ and $D_{j'}\{\alpha, \beta, \gamma\}$ of the rotation group. Their composition $D_j \times D_{j'}$ also gives us [73] a linear representation of the rotation group. We require to find the irreducible parts of which this representation is composed.

The objects of the linear representation D_j of order $(2j + 1)$ are

$$U_m = \frac{u_1^{j+m} u_2^{j-m}}{\sqrt{(j+m)!\,(j-m)!}} \quad (m = -j, -j+1, \ldots, j-1, j) \tag{142}$$

and those of $D_{j'}$ are

$$V_{m'} = \frac{v_1^{j'+m'} v_2^{j'-m'}}{\sqrt{(j'+m')!\,(j'-m')!}} \quad (m' = -j', -j'+1, \ldots, j'-1, j'), \tag{143}$$

where (u_1, u_2) and (v_1, v_2) undergo the same unitary transformations with $(+1)$ determinant [68]. If we form the $(2j + 1)(2j' + 1)$ quantities

$$W_{mm'} = U_m V_{m'} = \frac{u_1^{j+m} u_2^{j-m} v_1^{j'+m'} v_2^{j'-m'}}{\sqrt{(j+m)!\,(j-m)!\,(j'+m')!\,(j'-m')!}} \tag{144}$$

$$\begin{pmatrix} m = -j, \; -j+1 \; \ldots, \; j-1, j \\ m' = -j' -j'+1, \ldots, j'-1, j' \end{pmatrix},$$

these will be the objects in the linear representation of the rotation group defined by the composition $D_j \times D_{j'}$.

We shall assume in future that j and j' are either integers or half integers, i.e. to be more precise, we shall take linear representations of the unitary group in two variables with unity determinants.

Let k be an integer (or half an integer) satisfying the inequality

$$|j - j'| \leqslant k \leqslant j + j'. \tag{145}$$

We show that we can form $2k + 1$ linear combinations of magnitudes (144) such that they give a linear representation D_k of the rotation group.

For the proof, we form expressions of the type

$$L = (u_1 v_2 - u_2 v_1)^l (u_1 x_1 + u_2 x_2)^{2j-l} (v_1 x_1 + v_2 x_2)^{2j'-l}, \tag{146}$$

where l is a fixed integer satisfying the inequalities:

$$l \geqslant 0; \; l \leqslant 2j; \; l \leqslant 2j'. \tag{147}$$

If the variables (u_1, u_2) and (v_1, v_2) undergo the same linear transformation

$$u_1' = a_{11} u_1 + a_{12} u_2; \qquad v_1' = a_{11} v_1 + a_{12} v_2$$
$$u_2' = a_{21} u_1 + a_{22} u_2; \qquad v_2' = a_{21} v_1 + a_{22} v_2$$

with $(+1)$ determinant, i.e. $a_{11}a_{22} - a_{12}a_{21} = 1$, it may easily be seen that the first factor in (146) remains unchanged. For

$$u_1' v_2' - u_2' v_1' = (a_{11}a_{22} - a_{12}a_{21})(u_1 v_2 - u_2 v_1).$$

It is clear that (146) is a homogeneous polynomial in x_1 and x_2 of degree $2(j + j' - l)$. It therefore consists of terms of the form

$$a_s x_1^s x_2^{2(j+j'-l)-s} \qquad (s = 0, 1, \ldots, 2(j + j' - l))$$

On introducing the notation:

$$k = j + j' - l, \tag{148}$$

$$y_{m''} = \frac{x_1^{k+m''} x_2^{k-m''}}{\sqrt{(k+m'')!\,(k-m'')!}} \tag{149}$$

$$(m'' = -k, -k+1, \ldots, k-1, k),$$

we can write (146) as follows:

$$L = \sum_{m''=-k}^{+k} c_{m''} y_{m''}. \tag{150}$$

The coefficients $c_{m''}$ are dependent on the variables (u_1, u_2) and (v_1, v_2).

It follows at once from (146) that $c_{m''}$ is a homogeneous polynomial in (u_1, u_2) of degree $2j$ and a homogeneous polynomial in (v_1, v_2) of degree $2j'$, i.e. $c_{m''}$ will consist of terms of the form

$$a'_{pq} u_1^p u_2^{2j-p} v_1^q v_2^{2j'-q},$$

or we can say, on taking into account (142) and (143), that $c_{m''}$ is a linear combination of products:

$$c_{m''} = \sum_m \sum_{m'} d_{mm'}^{(m'')} U_m V_{m'} \; (m'' = -k, -k+1, \ldots, k-1, k), \tag{151}$$

where the coefficients $d_{mm'}^{(m'')}$ no longer contain u_k and v_k. We observe that, in (146), the variables u_1 and v_1 only appear either in association with the factor x_1, or in the first factor of (146), in which the sum of the indices of u_1 and v_1 is equal to l. On observing that $y_{m''}$ contains $x_1^{k+m''}$, we can say that, in the terms of (151), the sum of the indices of u_1 and v_1 is $k + m'' + l$, or by (148), the sum is $j + j' + m''$. But U_m contains u^{j+m} and $V_{m'}$ contains $v_1^{j'+m'}$, and hence it follows immediately that each of expressions (151) contains only the products $U_m V_{m'}$, for which $m + m' = m''$. We now show that linear combinations (151) of the $U_m V_{m'}$ in fact give a linear representation of the rotation group equivalent to D_k.

We first recall the definition of contragredient transformation. Given the two linear transformations

$$(x_1', \ldots, x_n') = A\,(x_1, \ldots, x_n) \text{ and } (y_1', \ldots, y_n') = B\,(y_1, \ldots, y_n),$$

the necessary and sufficient condition for the equation

$$x_1' y_1' + \ldots + x_n' y_n' = x_1 y_1 + \ldots + x_n y_n$$

to be valid is for B to be contragredient to A, i.e. $B = A^{(*)-1}$ (cf. [21] and [40]).

Let the variables (u_1, u_2) and (v_1, v_2) undergo a simultaneous unitary transformation A with $(+1)$ determinant. Suppose that the variables x_1 and x_2 have now undergone a transformation $A^{(*)-1}$ contragredient to A. It follows from the definition of contragredience that the sums

$$u_1 x_1 + u_2 x_2 \text{ and } v_1 x_1 + v_2 x_2$$

now remain invariant. As was proved above, the first factor in (146) also remains unchanged with the above transformation. The total sum L therefore remains unchanged, in other words, by (150), the variables $c_{m''}$ undergo a transformation B, contragredient to the transformation C suffered by the $y_{m''}$.

We bring in the new variables:

$$z_{m''} = \frac{u_1^{k+m''} u_2^{k-m''}}{\sqrt{(k+m'')!\,(k-m'')!}} \qquad (m'' = -k, -k+1, \ldots, k-1, k).$$

We can write on applying the binomial formula:

$$(u_1 x_1 + u_2 x_2)^{2k} = (2k)! \sum_{m''=-k}^{+k} z_{m''} y_{m''}.$$

The left-hand side remains unchanged by the transformations, and the same can therefore be said of the right-hand side, i.e. the variables $z_{m''}$ undergo the same transformation B, contragredient to C, as the variables $c_{m''}$. But we know that variables $z_{m''}$ in fact give us a linear representation D_k of the rotation group, if (u_1, u_2) are the objects of the unitary group with $(+1)$ determinants. Our assertion is therefore proved.

We can thus form $(2k+1)$ linear combinations of variables (144), which we shall interpret as vector components in space with $(2j+1)(2j'+1)$ dimensions, and the combinations give a linear representation D_k of the rotation group. On taking into account equation (148) and

inequality (147), the following values are seen to be assignable to the number k:

$$k = j + j', j + j' - 1, \ldots, |j - j'|. \tag{152}$$

We now find how many linear combinations of variables (144) can be formed. We assume for definiteness that $j \geqslant j'$. The total number of linear combinations will be

$$(2j + 2j' + 1) + (2j + 2j' - 1) + \ldots + (2j - 2j' + 1).$$

This is the sum of an arithmetic progression, the number of terms being

$$\frac{(2j + 2j' + 1) - (2j - 2j' + 1)}{2} + 1 = 2j' + 1,$$

and the total number of combinations is $(2j + 1)(2j' + 1)$, i.e. it is equal to the number of variables (144). The same result would be obtained on assuming $j < j'$. On writing for brevity:

$$(2j + 1)(2j' + 1) = r,$$

the linear combinations can be denoted by

$$w_1, w_2, \ldots, w_r, \tag{153}$$

on the assumption that the combinations run in the same order as the linear representations D_k, where k has the values given in (152). As a result of a unitary transformation with (+1) determinant on variables (u_1, u_2) and (v_1, v_2), we get new values $U'_m V'_{m'}$ of variables (144) and new values w'_s $(s = 1, 2, \ldots, r)$ of variables (153), where the w'_s are given in terms of the w_s by the quasi-diagonal matrix

$$[D_{j+j'}, D_{j+j'-1}, \ldots, D_{|j-j'|}], \tag{155}$$

and each D_k corresponds to the unitary transformation to which the (u_1, u_2) and (v_1, v_2) have been subjected. We show further that the linear forms (153) of magnitudes (144) are line arlyindependent. Let T be the matrix of the linear transformation with the aid of which the w_s are expressed in terms of the variables (144). The direct product $D_j \times D_{j'}$ is the matrix of the linear transformation for variables (144), and we have by the above:

$$(D_{j+j'}, D_{j+j'-1}, \ldots, D_{|j-j'|}) = T(D_j \times D_{j'})\,T^{-1}, \tag{155}$$

which gives the decomposition of the direct product into irreducible parts. The above expression is more usually written as follows:

$$D_j \times D_{j'} = D_{j+j'} + D_{j+j'-1} + \ldots + D_{|j-j'|}. \tag{156}$$

We recall that each D_k is defined by a unitary transformation and is to be written out in full as $D_k\begin{Bmatrix} a, & b \\ -\bar{b}, & \bar{a} \end{Bmatrix}$. The result obtained may be generalized for any number of factors. For instance, we can write

$$\begin{aligned} D_1 \times D_1 \times D_1 &= (D_2 + D_1 + D_0) \times D_1 = \\ &= D_3 + D_2 + D_1 + D_2 + D_1 + D_0 + D_1 = \\ &= D_3 + 2D_2 + 3D_1 + D_0 . \end{aligned}$$

D_1 itself is a third order matrix [68]. The direct product $D_1 \times D_1$ is a ninth order matrix, and finally $D_1 \times D_1 \times D_1$ is a matrix of order twenty-seven. The above equation shows that this last matrix is equivalent, with any choice of unitary transformation, to the diagonal matrix

$$[D_3, D_2, D_2, D_1, D_1, D_1, D_0] .$$

The order of this last matrix is [68]:

$$(2 \cdot 3 + 1) + 2(2 \cdot 2 + 1) + 3(2 \cdot 1 + 1) + (2 \cdot 0 + 1) = 27 .$$

We now prove the linear independence of the w_s, as linear forms of magnitudes (144). The w_s are the $c_{m''}$ in the previous notation, except that we have to remember that we can take different values of k, or what amounts to the same thing, different values of l, when forming the $c_{m''}$, so that it would be more correct to write $c_{m''}^{(l)}$. As we have seen above, each $c_{m''}^{(l)}$ is expressed solely in terms of the $U_m V_{m'}$, for which $m + m' = m''$. Hence it follows at once that only the $c_{m''}^{(l)}$ with different l but the same m'' can be linearly dependent. On removing the brackets of the last two factors in (146) and collecting terms in $x_1^{k+m''} x_2^{k-m''}$, where k is given by (148), we in fact obtain, up to a constant factor, the $c_{m''}^{(l)}$ in terms of u_k and v_k. They are clearly the products of $(u_1 v_2 - u_2 v_1)^l$ and a polynomial with positive integral coefficients in u_1, u_2, v_1 and v_2. It may readily be seen that these expressions cannot be linearly dependent with different l. Suppose, say, that we had linear dependence of the type:

$$a_1 c_{m''}^{(l_1)} + a_2 c_{m''}^{(l_2)} a_3 c_{m''}^{(l_3)} = 0,$$

where $l_1 < l_3 < l_3$ and the a_k are non-zero constants. This relationship must be satisfied as an identity for any u_1, u_2, v_1, v_2. Suppose, for instance, $u_2 = v_1 = v_2 = 1$. By what has been said about the form of the $c_{m''}^{(l)}$ we get a relationship of the type

$$a_1 (u_1 - 1)^{l_1} p_1(u_1) + a_2 (u_1 - 1)^{l_2} p_2(u_1) + a_3 (u_1 - 1)^{l_3} p_3(u_1) = 0,$$

where the $p_k(u_1)$ are polynomials in u_1 with positive integral coefficients. On dividing by $(u_1 - 1)^{l_1}$ then setting $u_1 = 1$, the above gives us $a_1 = 0$, which contradicts what has been said and thus proves the impossibility of a linear relationship.

Of course we could actually have constructed the expressions for the w_s in terms of variables (144) by removing the brackets in (146).

76. Orthogonality. Matrices forming non-equivalent unitary irreducible representations have the property generally known as *orthogonality*. They are often employed in applications of group theory to physics. We first of all formulate this property.

Let G be a finite group of order m with elements

$$G_1, G_2, \ldots, G_m$$

and let

$$A^{(1)}, \ldots, A^{(m)} \text{ and } B^{(1)}, \ldots, B^{(m)}$$

be two systems of matrices giving linear representations of G. If we write small letters with two subscripts for the matrix elements and assume that the representations are non-equivalent and irreducible and consist of unitary matrices, we shall find that we have the following equation:

$$\sum_{s=1}^{m} a_{ij}^{(s)} \overline{b_{kl}^{(s)}} = 0, \tag{157}$$

this being valid for any values of the subscripts. Similar equations apply for a single irreducible unitary representation. Let p be the order of matrices $A^{(s)}$, yielding an irreducible unitary representation. We have the following equations:

$$\sum_{s=1}^{m} a_{ij}^{(s)} \overline{a_{kl}^{(s)}} = \frac{m}{p} \delta_{ik} \delta_{jl}, \tag{158}$$

i.e. the sum on the left is zero if the pairs of numbers (i, j) and (k, l) are different, and is equal to m/p if the pairs are the same.

The proof of orthogonality is based on Theorem III of [66]. We first recall the multiplication rule in the case of rectangular (not square) matrices. Let C and D be matrices with elements

$$\{D\}_{ik} \begin{pmatrix} i = 1, 2, \ldots, n_1 \\ k = 1, 2, \ldots, n_2 \end{pmatrix} \quad \text{and} \quad \{C_{jl}\} \begin{pmatrix} j = 1, 2, \ldots, n_2 \\ l = 1, 2, \ldots, n_3 \end{pmatrix},$$

the number n_2 of columns of D being the same as the number of rows of C. The elements of the product DC are defined by the usual expression

$$\{DC\}_{ik} = \sum_{s=1}^{n_2} \{D\}_{is} \{C\}_{sk}.$$

The new matrix DC will clearly have n_1 rows and n_3 columns.

We now state a fundamental theorem.

THEOREM. *If unitary matrices $A^{(s)}$ of order p and unitary matrices $B^{(s)}$ of order q give non-equivalent irreducible representations of a group G, and if a rectangular matrix C with p rows and q columns satisfies for all s:*

$$A^{(s)} C = C B^{(s)} \qquad (s = 1, 2, \ldots, m), \tag{159}$$

C is a zero matrix, i.e. all its elements are zero.

We first take the case $p = q$, when C is a square matrix. If the determinant of C differs from zero, there exists C^{-1}, and it follows from (159) that

$$A^{(s)} = C B^{(s)} C^{-1},$$

i.e. the two representations are equivalent, which contradicts the hypothesis of the theorem. The determinant of C must therefore vanish. Suppose that not all the elements of C are zero and that we write them as c_{ik}. We know that the linear forms

$$c_{i1} x_1 + \ldots + c_{ip} x_p \qquad (i = 1, 2, \ldots, p)$$

define with arbitrary x_s a subspace with a number of dimensions equal to the rank of C [14], i.e. the subspace here has a number of dimensions $\geqq 1$ and $< p$. In other words, we are concerned here with a subspace R and not the total space of p dimensions. We write (159) as a linear transformation on a vector with components $(x_1, \ldots, x_p)$:

$$A^{(s)} C (x_1, \ldots, x_p) = C B^{(s)} (x_1, \ldots, x_p) \qquad (s = 1, 2, \ldots, m).$$

The $C(x_1, \ldots, x_p)$ on the left is an arbitrary vector of R, whilst the whole of the right-hand side, representing a linear transformation C on a vector $B^{(s)}(x_1, \ldots, x_p)$, also belongs to R. In other words, the transformation $A^{(s)}$ on any vector of R again yields a vector of R. In this case, as we know from [66], the $A^{(s)}$ give a reducible representation, which contradicts the hypothesis of the theorem.

This proof remains in force if $p > q$. The rank of C is now always less than p, and the linear forms

$$c_{i1} x_1 + \ldots + c_{iq} x_q \qquad (i = 1, 2. \ldots, p)$$

define a subspace R with a number of dimensions less than p; thus the proof remains as before. Suppose finally that $p < q$; we pass to the transposed matrices in (159), which gives us

$$B^{(s)(*)} C^{(*)} + C^{(*)} A^{(s)(*)}.$$

The order q of $B^{(s)(*)}$ is higher than the order p of $A^{(s)(*)}$, and we conclude from this as above that the unitary matrices $B^{(s)(*)}$ leave a subspace unchanged, so that we can reduce them to the quasi-diagonal form by a suitable choice of fundamental vectors. The matrices $B^{(s)}$ will also become quasi-diagonal, which contradicts the hypothesis of the theorem. The theorem is thus proved.

We could have omitted the condition in the theorem that $A^{(s)}$ and $B^{(s)}$ are unitary. As we know, these can always be assumed unitary if we are prepared to pass to similar representations, in which case we get a new matrix C_1 instead of C in (159), C_1 being connected with C by a relationship of the form

$$C = D_1 C_1 D_2;$$

and since C_1 is the null matrix, the same can be said of C.

We now turn to the proof of (157). We introduce the notations $A(G_s)$ and $B(G_s)$ instead of $A^{(s)}$ and $B^{(s)}$, where G_s is the element of G to which $A^{(s)}$ and $B^{(s)}$ correspond. Let X be any matrix with p rows and q columns. We introduce the matrix

$$C = \sum_{s=1}^{m} A(G_s)\,XB(G_s)^{-1} \tag{160}$$

and show that it satisfies (159).

Let G_t be a fixed element of G. We have

$$A(G_t)\,C = \sum_{s=1}^{m} A(G_t)\,A(G_s)\,XB(G_s)^{-1}.$$

But by the definition of linear representation:

$$A(G_t)\,A\,G_s = A(G_t\,G_s) \quad \text{and} \quad B(G_t)\,B(G_s) = B(G_t\,G_s),$$

and hence

$$A(G_t)\,C = \sum_{s=1}^{m} A(G_t\,G_s)\,XB(G_t\,G_s)^{-1}\,B(G_t).$$

If G_s runs over all elements of the group, the same can be said of the product $G_t\,G_s$, so that we can write the equation above as

$$A(G_t)\,C = CB(G_t),$$

i.e. the matrix C defined by (160) in fact satisfies (159), and C is consequently a null matrix. We thus have, for any choice of matrix X:

$$\sum_{s=1}^{m} A(G_s)\,XB(G_s)^{-1} = 0.$$

Suppose that a fixed element $\{X\}_{jl}$ of X is unity, and the remainder zero. The last equation now gives us

$$\sum_{s=1}^{m} \{A(G_s)\}_{ij}\,\{B(G_s)^{-1}\}_{lk} = 0.$$

Since the matrices are unitary, $B(G_s)$ is obtained from $B(G_r)^{-1}$ by replacing rows by columns and all the elements by their conjugates, so that the last equation becomes in the previous notation:

$$\sum_{s=1}^{m} a_{ij}^{(s)}\,\overline{b_{kl}^{(s)}} = 0,$$

which is the same as (157).

Similarly, by constructing the matrix

$$D = \sum_{s=1}^{m} A(G_s)\,XA(G_s)^{-1},$$

where X is any square matrix of order p, we can show that

$$A(G_s)D = DA(G_s) \quad (s = 1, 2, \ldots, m),$$

and we can say from Theorem III of [66] that D is scalar matrix, or

$$\sum_{s=1}^{m} A(G_s)XA(G_s)^{-1} = cI,$$

where the number c depends on the choice of X. Again, let $\{X\}_{jl} = 1$ and the remaining elements of X be zero, and let c_{jl} denote the corresponding value of the number c. We can write

$$\sum_{s=1}^{m} \{A(G_s)\}_{ij}\{A(G_s)^{-1}\}_{lk} = c_{jl}\,\delta_{ik}. \tag{161}$$

To find c_{jl}, we put $i = k$ and sum over i from 1 to p:

$$pc_{jl} = \sum_{s=1}^{m}\sum_{i=1}^{p} \{A(G_s)^{-1}\}_{li}\{A(G_s)\}_{ij} = \sum_{s=1}^{m} \{I\}_{lj}.$$

If $l = j$, the right-hand side is equal to m, whilst it vanishes with $l \neq j$. Hence $c_{jl} = (m/p)\,\delta_{jl}$, and (161) can therefore be re-written:

$$\sum_{s=1}^{m} \{A(G_s)\}_{ij}\{A(G_s)^{-1}\}_{lk} = \frac{m}{p}\,\delta_{ik}\,\delta_{jl}, \tag{162}$$

which is the same as (158) if we take into account the fact that $A(G_s)$ is unitary.

Relationship (157) may easily be seen to hold, not merely for unitary, but for arbitrary non-equivalent and irreducible group representations. Let $A'(G_s)$ and $B'(G_s)$ be two such representations of degrees p and q, whilst $A(G_s)$ and $B(G_s)$ are unitary representations equivalent to them, so that

$$A(G_s) = C_1 A'(G_s) C_1^{-1}; \quad B(G_s) = C_2 B'(G_s) C_2^{-1},$$

where C_1 and C_2 are definite matrices not depending on s. We have by the unitariness of $B(G_s)$:

$$B(G_s)^{-1} = \overline{B(G_s)^*} = \overline{(C_2^{-1})^*}\,\overline{B'(G_s)^*}\,\overline{C_2^*},$$

and (157) can be written as

$$\sum_{s=1}^{m} C_1 A'(G_s) C_1^{-1} X \overline{(C_2^{-1})^*}\,\overline{B'(G_s)^*}\,\overline{C_2^*} = 0,$$

whence, on multiplying on the left by C_1^{-1} and on the right by $\overline{(C^*)^{-1}}$, and introducing the arbitrary matrix $Y = C_1^{-1} X \overline{(C_2^{-1})^*}$ with p rows and q columns, we get

$$\sum_{s=1}^{m} A'(G_s)\,Y\,\overline{B'(G_s)^*} = 0,$$

and therefore, using the arbitrariness of Y as above:

$$\sum_{s=1}^{m} a_{ij}^{(s)} \overline{b_{kl}^{(s)}} = 0.$$

We notice also that (162) is valid for any representation, unitary or not, as follows from the proof and the fact that it is not necessary to mention the unitariness of the $A^{(s)}$ and $B^{(s)}$ in the statement of the previous theorem.

77. Characters. Suppose, as above, that $A(G_s)$, $B(G_s)$ are two non-equivalent irreducible representations of orders p, q of a group G with elements $G_1, G_2, \ldots, G_m$. We shall write $X(G_s)$, $X'(G_s)$ for the traces of the matrices of the representations, i.e. the sums of their diagonal elements:

$$X(G_s) = \sum_{i=1}^{p} \{A(G_s)\}_{ii}; \quad X'(G_s) = \sum_{k=1}^{q} \{B(G_s)\}_{kk}.$$

These numbers are known as the *characters* of the representations. The characters of equivalent representations are clearly the same [27]; also, we can assume that the representations in question are unitary. The orthogonality formula gives

$$\sum_{s=1}^{m} \{A(G_s)\}_{ii} \overline{\{B(G_s)\}_{kk}} = 0,$$

and summation over i and k gives the orthogonality formula for the characters:

$$\sum_{s=1}^{m} X(G_s) \overline{X'(G_s)} = 0. \tag{163}$$

Similarly, (158) gives

$$\sum_{s=1}^{m} \{A(G_s)\}_{ii} \overline{\{A(G_s)_{kk}\}} = \frac{m}{p} \delta_{ik}$$

and we have from summation over i and k:

$$\sum_{s=1}^{m} X(G_s) \overline{X(G_s)} = m. \tag{164}$$

We shall prove a number of theorems by using these formulae.

THEOREM 1. *The necessary and sufficient condition for two irreducible representations to be equivalent is that their characters are the same.*

We have already mentioned that the characters of equivalent (reducible or irreducible) representations are the same, so that the necessity of the condition is established. We now assume the converse, that the systems of characters of two irreducible representations are the same, i.e. $X(G_s) = X'(G_s)$ ($s = 1, 2, \ldots, m$), and prove the equivalence of the representations. We have by (164):

$$\sum_{s=1}^{m} X(G_s) \overline{X'(G_s)} = m,$$

whence the equivalence follows, since otherwise we should have relationship (163). We notice the obvious point that the matrices in equivalent representations must be of the same order. Corresponding to each irreducible representation, we introduce vectors in the complex m-dimensional space R_m with components:

$$\frac{1}{\sqrt{m}} X(G_1)\,, \quad \frac{1}{\sqrt{m}} X(G_2), \ldots, \quad \frac{1}{\sqrt{m}} X(G_m)\,.$$

These vectors are normalized by virtue of (164), and vectors corresponding to non-equivalent representations are mutually orthogonal by (163). Hence it follows that *there cannot exist more than m non-equivalent irreducible representations of a group G of order m*. We shall later define more precisely the total number of non-equivalent irreducible representations of a group. We shall denote this number by the letter l for the present. Let $\omega^{(i)}$ denote these non-equivalent irreducible representations ($i = 1, 2, \ldots, l$) and let their characters be

$$X^{(i)}(G_1), \quad X^{(i)}(G_2), \ldots, X^{(i)}(G_m) \qquad (i = 1,\ 2, \ldots,\ l)\,.$$

Suppose that there exists a representation ω with characters

$$X(G_1), \quad X(G_2), \ldots, \ X(G_m)\,.$$

As a result of reduction, ω is given by quasi-diagonal matrices formed from the matrices of representations $\omega^{(i)}$. We thus have for the characters:

$$X(G_s) = \sum_{i=1}^{l} a_i X^{(i)}(G_s)\,, \tag{165}$$

where the a_i are non-negative integers which show us how many times the representation $\omega^{(i)}$ appears in the constitution of ω after its reduction.

Expressions can be derived for the coefficients a_i in terms of the characters of representation ω. Let k be one of the numbers $1, 2, \ldots, l$. We multiply both sides of (165) by $\overline{X^{(k)}(G_s)}$ and sum over s. We obtain, using (163) and (164):

$$\sum_{s=1}^{m} X(G_s)\, \overline{X^{(k)}(G_s)} = a_k m\,,$$

whence

$$a_k = \frac{1}{m} \sum_{s=1}^{m} X(G_s)\, \overline{X^{(k)}(G_s)}\,. \tag{166}$$

This expression yields a definite value for each a_k, whence we get the following theorem.

THEOREM 2. *Every reducible representation decomposes into a unique set of irreducible representations.*

By using (166), we can easily generalize Theorem 1 for the case of any representations, irreducible or not.

THEOREM 3. *The necessary and sufficient condition for two representations to be equivalent is for their characters to be the same.*

The necessity of the condition has been noted in the proof of Theorem 1. Conversely, if the characters $X(G_s)$ of two representations are the same, we get

like values for the a_k by (166), and both representations consequently reduce to a quasi-diagonal matrix composed of the same irreducible representations. We can assume here, on passing to an equivalent representation if necessary, that the irreducible representations in question are arranged in the quasi-diagonal matrix in the same order, since a permutation of rows and columns is equivalent to passage to equivalent representation.

Representations with the same characters thus reduce to the same quasi-diagonal matrix, i.e. they are equivalent.

We now turn to the investigation of the total number l of irreducible, non-equivalent representations of the group G. The group elements are distributed into classes. We find in the same class the elements obtained from one of them G_t with the aid of the expression:

$$G_s G_t G_s^{-1} \qquad (s = 1, 2, \ldots, m)\,.$$

Similar matrices with the same trace correspond to all these elements in any representation. Let r be the number of classes in G. By what has been said above, every linear representation of G has not more than r different characters, where each character corresponds, not to individual elements, but to all the elements of a given class. Let the class C_1 consist of g_1 elements, C_2 of g_2 elements, and finally, C_r of g_r elements. The terms of sum (163) are the same for elements of the same class, and on writing $X(C_k)$, $X'(C_k)$ for the characters corresponding to the elements of class C_k, we can re-write (163) for two non-equivalent irreducible representations as

$$\sum_{k=1}^{r} X(C_k)\,\overline{X'(C_k)}\,g_k = 0\,,$$

whilst (164) becomes

$$\sum_{k=1}^{r} X(C_k)\,\overline{X(C_k)}\,g_k = m\,.$$

We thus have, for the characters $X^{(i)}(C_k)$ of non-equivalent irreducible representations $\omega^{(i)}$ $(i = 1, 2, \ldots, l)$:

$$\sum_{k=1}^{r} X^{(i_1)}(C_k)\,\overline{X^{(i_2)}(C_k)}\,g_k = 0 \quad \text{for } i_1 \neq i_2$$
$$\sum_{k=1}^{r} X^{(i)}(C_k)\,\overline{X^{(i)}(C_k)}\,g_k = m\,. \tag{167}$$

We introduce into the r-dimensional space R_r l vectors with components

$$\sqrt{\frac{g_1}{m}}\,X^{(i)}(C_1), \quad \sqrt{\frac{g_2}{m}}\,X^{(i)}(C_2), \ldots, \quad \sqrt{\frac{g_r}{m}}\,X^{(i)}(C_r) \qquad (i = 1, 2, \ldots, l)\,.$$

The above equations show that these vectors are mutually orthogonal and normalized, and consequently linearly independent. It follows that the number l of them is not greater than the number of dimensions, i.e. $l \leqslant r$. This gives us the theorem:

THEOREM 4. *The number of non-equivalent, irreducible representations of a group is not greater than the number of classes of the group.*

It is shown in the next article that we always have $l = r$. Since we have just proved that $l \leqslant r$, the equality follows if we can show that $l \geqslant r$. The proof of this latter inequality is bound up with the introduction of certain new concepts and relationships regarding characters which are of interest in themselves.

We establish a further relationship between the characters of any irreducible representation. Let the class C_k consists of the elements $G_1^{(k)}, G_2^{(k)}, \ldots, G_{g_k}^{(k)}$. The expression $G_s G_i^{(k)} G_s^{-1}$ $(i = 1, 2, \ldots, g_k)$, where G_s is any element of the group, again gives us all the elements of class C_k, though now in a different order. It follows that, if we take the set of all the products of elements of two classes C_p and C_q:

$$G_u^{(p)} G_v^{(q)} \quad (u = 1, 2, \ldots, g_p; \quad v = 1, 2, \ldots, g_q)\,, \tag{168}$$

the expression

$$G_s\, G_u^{(p)}\, G_v^{(q)}\, G_s^{-1} = (G_s\, G_u^{(p)} G_s^{-1})\, (G_s\, G_v^{(q)}\, G_s^{-1})$$

gives us the same set of elements. This implies the following property of set (168): if an element belongs to the set, the whole of the class containing the element likewise belongs to the set, where each element of the class appears in set (168) the same number of times. *We write a_{pqk} for the non-negative integer indicating how many times elements of class C_k appear in set* (168). All this may be expressed in a purely symbolic form as

$$C_p C_q = \sum_{k=1}^{r} a_{pqk} C_k \tag{169}$$

or

$$(G_1^{(p)} + G_2^{(p)} + \cdots + G_{g_p}^{(p)})\, (G_1^{(q)} + G_2^{(q)} + \cdots + G_{g_q}^{(q)}) =$$

$$= \sum_{k=1}^{r} a_{pqk}\, (G_1^{(k)} + G_2^{(k)} + \cdots + G_{g_k}^{(k)})\,. \tag{170}$$

Let $A(G_s)$ be the nth order matrices of an irreducible linear representation of group G. We form the sum of the matrices corresponding to elements of class C_k and call the new matrix $A(C_k)$:

$$A(C_k) = \sum_{j=1}^{g_k} A(G_j^{(k)})\,.$$

On observing that the elements $G_s\, G_i^{(k)} G_s^{-1}$ with $i = 1, 2, \ldots, g_k$ and any G_s of G give the total set of elements of class C_k, it will be seen that the matrix $A(C_k)$ commutes with all the matrices $A(G_s)$. Hence it follows that $A(C_k)$ is a scalar matrix [66], so that we can write:

$$A(C_k) = b_k I \qquad (k = 1, 2, \ldots, r)\,, \tag{171}$$

where the b_k are numbers. If we make use of the definition of the numbers a_{pqk}, i.e. of symbolic expression (170), we get the following relationship between the b_k:

$$b_p\, b_q = \sum_{k=1}^{r} a_{pqk}\, b_k\,. \tag{172}$$

The trace of matrix $A(C_k)$ is equal to the sum of the traces of the $A(G_i^{(k)})$ $(i = 1, 2, \ldots, g_k)$, i.e. is equal to $g_k X(C_k)$. On the other hand, it follows from (171) that the trace of $A(C_k)$ is equal to nb_k, i.e. $nb_k = g_k X(C_k)$, whence

$$b_k = \frac{g_k}{n} X(C_k)\,,$$

and relationship (172) leads to the following theorem.

THEOREM 5. *The following relationships hold between the characters of any irreducible representation formed by n-th order matrices:*

$$g_p X(C_p)\, g_q X(C_q) = n \sum_{k=1}^{r} a_{pqk} X(C_k)\,. \tag{173}$$

We note that one of the classes C_k is that consisting merely of the identity element E of the group G. Corresponding to this we always have the unit matrix of a linear representation, the trace of which is equal to its order n. This class will always be denoted by C_1, so that $X(C_1) = n$, and the above expression can be re-written:

$$g_p X(C_p)\, g_q X(C_q) = X(C_1) \sum_{k=1}^{r} a_{pqk}\, g_k X(C_k)\,. \tag{174}$$

We now find the values of constants a_{pq1}. There corresponds to each class C_p a class $C_{p'}$, consisting of the inverse elements to those of C_p. This follows at once from the definition of class and the fact that the equation $G_s G_t G_s^{-1} = G_u$ leads to $G_s G_t^{-1} G_t^{-1} = G_u^{-1}$.

The class $C_{p'}$ can coincide with C_p, i.e. it may happen that $p = p'$. In every case, C_p and $C_{p'}$ contain the same number of elements, i.e. $g_{p'} = g_p$. If we take $q = p'$ in (173) or (174), class C_1 will appear g_p times on the right-hand side, whilst with $q \neq p'$, the right-hand side does not contain C_1, i.e.

$$a_{pq1} = \begin{cases} 0 & \text{for } q \neq p'\,, \\ g_p & \text{for } q = p'\,. \end{cases} \tag{175}$$

78. Regular representations of groups. We have already mentioned a method of representing a finite group with the aid of a permutation group. Any permutation group can be expressed in the form of a transformation group.

For suppose we have the permutation

$$\begin{matrix} 1, & 2, & 3, & 4 \\ 2, & 4, & 3, & 1, \end{matrix}$$

this can be written as the linear transformation by which x_1 becomes y_2, x_2 becomes y_4, x_3 becomes y_3, and x_4 becomes y_1:

$$\begin{aligned} y_1 &= 0x_1 + 0x_2 + 0x_3 + x_4 \\ y_2 &= x_1 \;\; + 0x_2 + 0x_3 + 0x_4 \\ y_3 &= 0x_1 + 0x_2 + x_3 \;\; + 0x_4 \\ y_4 &= 0x_1 + x_2 \;\; + 0x_3 + 0x_4\,. \end{aligned}$$

We consider the following representation of a group G by a permutation group. We multiply the elements $G_1, G_2, \ldots, G_m$ on the right by an element G_s. This leads to a permutation of the elements, i.e. by what has been said above, to a matrix P_s, which is regarded as corresponding to the element G_s. This is generally known as a *regular representation of the group G*. One of the G_k is the identity element of the group, which we denote as usual by E. The unit matrix P_s corresponds to this, and its trace is therefore m, i.e. $X(E) = m$. On multiplication of elements $G_1, G_2, \ldots, G_m$ by some element G_s, no element G_k remains in place, i.e. all the diagonal elements are zero in the corresponding matrix, and in a regular representation $X(G_s) = 0$ for $G_s \neq E$.

Suppose that, on reduction, a regular representation contains the representation $\omega^{(k)}$, that we have discussed above, h_k times. We have with this, by what has already been said:

$$\sum_{t=1}^{l} h_t\, X^{(t)}(G_s) = \begin{cases} 0 & \text{for } G_s \neq E \\ m & \text{for } G_s = E\,. \end{cases} \tag{176}$$

On multiplying both sides of this equation by $\overline{X^{(k)}(G_s)}$ and summing over s, we get by (163) and (164):

$$h_k m = mX^{(k)}(E);$$

but if we write n_k for the order of the matrices in representation $\omega^{(k)}$, we have $X^{(k)}(E) = n_k$, whilst from above, $\overline{X^{(k)}(E)} = X^{(k)}(E) = h_k$, whence $n_k = h_k$, and (176) can be written as

$$\sum_{t=1}^{l} X^{(t)}(E)\, X^{(t)}(G_s) = \sum_{t=1}^{l} n_t\, X^{(t)}(G_s) = \begin{cases} 0 & \text{for } G_s \neq E \\ m & \text{for } G_s = E\,. \end{cases} \tag{177}$$

We thus arrive at the following theorem.

THEOREM 6. *A regular representation contains each irreducible representation $\omega^{(k)}$ a number of times equal to the order n_k of the matrices in the $\omega^{(k)}$, the characters of the $\omega^{(k)}$ being given by* (177).

We now write down (174) for representation $\omega^{(k)}$:

$$g_p\, X^{(t)}(C_p)\, g_q\, X^{(t)}(C_q) = X^{(t)}(C_1) \sum_{k=1}^{r} a_{pqk}\, X^{(t)}(C_k)\, \varsigma_k\,.$$

and we sum over t from 1 to l:

$$g_p\, g_q \sum_{t=1}^{l} X^{(t)}(C_p)\, X^{(t)}(C_q) = \sum_{k=1}^{r} a_{pqk} \sum_{t=1}^{l} X^{(t)}(C_1)\, X^{(t)}(C_k)\, \varsigma_k\,.$$

We obtain, on taking (177) into account:

$$g_p\, g_q \sum_{t=1}^{l} X^{(t)}(C_p)\, X^{(t)}(C_q) = a_{pq1}\, m\,,$$

i.e. by (175):

$$\sum_{t=1}^{l} X^{(t)}(C_p)\, X^{(t)}(C_q) = \begin{cases} 0 & \text{for } q \neq p' \\ \dfrac{m}{g_{p'}} & \text{for } q = p'\,. \end{cases} \tag{178}$$

We form the set of l homogeneous linear equations in $x_1, x_2, \ldots, x_r$:

$$\sum_{q=1}^{r} x_q X^{(k)}(C_q) = 0 \qquad (k = 1, 2, \ldots, l) \tag{179}$$

and show that it only has the zero solution.

For, on multiplying both sides of (179) by $\overline{X^{(k)}(C_p)}$ and summing over k, we get $x_{p'} = 0$, where p' is any of the numbers $1, 2, \ldots, r$. Since system (179) has only the zero solution, the number of equations is not less than the number of unknowns, i.e. $l \geqslant r$. We showed earlier that $l \leqslant r$, whence it follows that $l = r$, i.e.

THEOREM 7. *The total number of non-equivalent irreducible representations of a finite group G is equal to the number of classes of G.*

We further notice a consequence of Theorem 6. The regular representation of group G consists of matrices of order m. On the other hand, by Theorem 6, it contains each representation $\omega^{(k)}$, consisting of matrices of order n_k, n_k times.

This gives us the equation:

$$\sum_{k=1}^{r} n_k^2 = m, \tag{180}$$

which may be stated in words as:

THEOREM 8. *The sum of the squares of the orders of the non-equivalent irreducible representations $\omega^{(k)}$ is equal to the order of the group G.*

79. Examples of representations of finite groups. 1. We take the Abelian group G consisting of elements $A_2^k A_1^i$, where $i = 0, 1, 2, \ldots, m-1$; $k = 0, 1, 2, \ldots, n-1$, the A_1 and A_2 elements commute, $A_1^m = E$, $A_2^n = E$, and with $i = k = 0$ we have to take $A_2^0 A_1^0 = E$. Each individual element of G forms a class, and all the irreducible representations of the group are of the first degree. Let α and β be values of the mth and nth roots of unity. We associate with the element $A_2^k A_1^i$ the number $\beta^k \alpha^i$, and thus obtain a group representation, as is easily seen. Assigning to α and β all possible values of the above-mentioned roots, we get altogether mn different first degree representations. The total number of classes, i.e. elements, is also equal to mn, and all the non-equivalent irreducible representations are thus obtained. The construction of the representations is similar in the case when the number of factor "elements" (i.e. of A_s) is more than two.

2. We turn to the nth order dihedral group. It consists of the $2n$ elements:

$$E,\ A^i,\ T,\ TA^i \qquad (i = 1, 2, \ldots, n-1),$$

where

$$A^n = E; \quad T^2 = E; \quad TAT^{-1} = A^{-1} \quad (T^{-1} = T). \tag{181}$$

The last of the relationships written is immediately obvious from the geometrical meaning of rotations A and T. An immediate consequence of this is the relationship $TA^i T^{-1} = A^{-i}$. First let $n = 2m + 1$ be odd. The group will now consist of $(m + 2)$ classes. One of these contains E; m consist of the two elements A^s and A^{-s} $(s = 1, 2, \ldots, m)$; and one contains all the elements of the type

T and TA^i. All this may easily be verified with the aid of the above relationships.

There exist two first degree representations; in one, the number 1 is associated with each element; in the other, 1 is associated with element A and (-1) with T. Now let $\varepsilon = \cos 2\pi/n + i \sin 2\pi/n$. We can form m second order representations by associating with the elements A and T the matrices:

$$A \to \left\| \begin{matrix} \varepsilon^s, & 0 \\ 0, & \varepsilon^{-s} \end{matrix} \right\|; \quad T \to \left\| \begin{matrix} 0, & 1 \\ 1, & 0 \end{matrix} \right\| \quad (s = 1, 2, \ldots, m). \tag{182}$$

These matrices satisfy relationships (181) and in fact yield the group representation, since every relationship between elements A and T is a consequence of (181). The irreducibility of each of the representations follows from the fact that, otherwise, a representation would reduce to two first order representations, and the matrices of the representation would have to commute, which is not in fact the case for any s, as may easily be verified.

The non-equivalence of representations (182) for different s follows from the fact that the matrices corresponding to element A have different sets of characteristic roots ε^s and ε^{-s} for different s. All $(m + 2)$ non-equivalent, irreducible representations have thus been obtained. Equation (180) amounts in the present example to

$$2 \cdot 1^2 + m \cdot 2^2 = 4m + 2 = 2n.$$

With even $n = 2m$, representation (182) corresponding to the value $s = m$ has the form

$$A \to \left\| \begin{matrix} -1, & 0 \\ 0, & -1 \end{matrix} \right\|; \quad T \to \left\| \begin{matrix} 0, & 1 \\ 1, & 0 \end{matrix} \right\|$$

and splits up into the two first degree representations

$$A \to (-1); \quad T \to (+1) \quad \text{and} \quad A \to (-1); \quad T \to (-1).$$

To obtain this in addition, it is sufficient to utilize a matrix S such that STS^{-1} reduces to the diagonal form, the characteristic roots of T being clearly equal to ± 1. Thus with $n = 2m$, there are four first degree representations and $(m - 1)$ second degree. Equation (180) becomes

$$4 \cdot 1^2 + (m - 1)\, 2^2 = 4m = 2n.$$

3. We consider the representations of the tetrahedral group or, what amounts to the same thing, the alternating group isomorphic to it with $n = 4$ [59]. The group consists of four classes and its order is twelve. It must have four non-equivalent irreducible representations. The degrees of these representations must satisfy the equation

$$n_1^2 + n_2^2 + n_3^2 + n_4^2 = 12.$$

This equation has a unique positive integral solution, discounting the order of the terms on the left-hand side:

$$n_1 = n_2 = n_3 = 1; \quad n_4 = 3,$$

i.e. the group has three representations of the first degree and one of the third. In the first degree representations, the same number corresponds to elements of the same class, and the correspondences may easily be seen to be as follows for these three representations:

$$\begin{array}{llll} \mathrm{I} \to 1; & \mathrm{II} \to 1; & \mathrm{III} \to 1; & \mathrm{IV} \to 1 \\ \mathrm{I} \to 1; & \mathrm{II} \to 1; & \mathrm{III} \to \varepsilon; & \mathrm{IV} \to \varepsilon^2 \\ \mathrm{I} \to 1; & \mathrm{II} \to 1; & \mathrm{III} \to \varepsilon^2; & \mathrm{IV} \to \varepsilon, \end{array}$$

where

$$\varepsilon = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}\,.$$

The third order irreducible representation gives the tetrahedral group itself, i.e. the group of rotations of space (third order matrices) for which the tetrahedron is displaced into itself. If this representation were reducible, it would have to reduce to three first degree representations, which is impossible since the group is not Abelian. The last sections have been concerned with the theory for finite groups. An extension to rotation groups requires a more detailed treatment of infinite groups that depend on parameters. Before we pass to the general treatment of these latter groups, we consider the problem of linear representations of the Lorentz group. These representations, together with those of the rotation group, will serve us as fundamental examples of infinite parametrically dependent groups.

80. Representations of a linear group in two variables. We constructed in [68] linear representations of a unitary group in two variables, which led us to linear representations of rotation groups. Representations can similarly be constructed of a linear group in two variables with unity determinant:

$$\begin{aligned} x_1' &= ax_1 + bx_2 \\ x_2' &= cx_1 + dx_2 \end{aligned} \quad ad - bc = 1\,. \tag{183}$$

This leads us, by what was said in [64], to one and two-valued representations of the group of positive Lorentz transformations. We arrive at results entirely different to those of [68].

One possible linear representation of unitary group (93) is the representation by the group itself, i.e. the linear representation in which, corresponding to a given transformation, we have the same transformation. Another linear representation is easily seen to be the following: to each transformation (93) there corresponds the transformation with complex conjugate coefficients:

$$y_1' = \bar{a}y_1 + \bar{b}y_2; \quad y_2' = -by_1 + ay_2\,.$$

But this representation is equivalent to the previous one, as follows directly from the fairly obvious equation:

$$\begin{Vmatrix} 0, 1 \\ -1, 0, \end{Vmatrix} \begin{Vmatrix} a, b \\ -\bar{b}, \bar{a} \end{Vmatrix} = \begin{Vmatrix} \bar{a}, \bar{b} \\ -b, a \end{Vmatrix} \begin{Vmatrix} 0, 1 \\ -1, 0 \end{Vmatrix}.$$

The conjugate representation for group (183):

$$y_1' = \bar{a}y_1 + \bar{b}y_2; \quad y_2' = \bar{c}y_1 + \bar{d}y_2 \tag{184}$$

is not equivalent to group (183) itself. To see this, we need only consider the case $b = c = 0$. The matrix transformation of (183) now has characteristic roots a and d, whilst the matrix of (184) has roots $\bar{a}$ and $\bar{d}$. We can clearly choose complex numbers a and d satisfying the condition $ad = 1$ such that the set $\bar{a}$ and $\bar{d}$ is different from the set a and d, so that the corresponding transformations cannot be similar. We have thus already obtained two non-equivalent second degree representations — the group itself (183) and group (184). We discuss below the irreducibility of the representations.

We can moreover construct representations of group (183) precisely as we did in [68]. We only need to replace $\bar{a}$ by d and $\bar{b}$ by $(-c)$. This gives us the following representation of order $(2j + 1)$, where j is a non-negative integer or half integer:

$$D_j \begin{Bmatrix} a, b \\ c, d \end{Bmatrix}_{ls} = \sum_k \frac{\sqrt{(j+l)!\,(j-l)!\,(j+s)!\,(j-s)!}}{k!\,(j-k-s)!\,(j+l-k)!\,(k+s-l)!} \times$$
$$\times a^{j+l-k} b^k c^{k+s-l} d^{j-k-s} \left(j = 0, \frac{1}{2}, \ 1, \ \frac{3}{2}, \ \ldots\right). \tag{185}$$

Here l and s run over the following values:

$$l \text{ and } s = -j, -j+1, \ldots, j-1, j,$$

and the summation over k is defined by the inequalities:

$$k \geqslant 0; \ k \geqslant l - s; \ k \leqslant j - s; \ k \leqslant j + l.$$

We have to take $0! = 1$ and $0^0 = 1$ in (185). The identity representation by unity is obtained with $j = 0$. We can at once write further representations in addition to (185) by replacing the numbers a, b, c, d by their conjugates on the right-hand side of (185). We shall denote the corresponding representations as follows:

$$\overline{D}_{j'} \begin{Bmatrix} a, b \\ c, d \end{Bmatrix} \quad \left(j' = 0, \ \frac{1}{2}, \ 1, \ \frac{3}{2}, \ \ldots\right). \tag{186}$$

We can now form a composition of representations (185) and (186)

[73], as a result of which a new representation of order $(2j+1)(2j'+1)$ is obtained. We denote this as follows:

$$E_{j,j'}\begin{Bmatrix} a, b \\ c, d \end{Bmatrix}. \tag{187}$$

By using (185), we can easily write down the elements of the matrices corresponding to this representation. We take two different representations (187), though of the same order:

$$E_{p,q}\begin{Bmatrix} a, b \\ c, d \end{Bmatrix} \text{ and } E_{p_1,q_1}\begin{Bmatrix} a, b \\ c, d \end{Bmatrix}; \ (2p+1)(2q+1) = (2p_1+1)(2q_1+1).$$

We show that these two representations are not equivalent. We put $b = c = 0$. Matrices (185) now reduce to the diagonal form with diagonal elements

$$D_j\begin{Bmatrix} a, 0 \\ 0, d \end{Bmatrix}_{ll} = a^{j+l}d^{j-l} \quad (l = -j, -j+1, \ldots, j-1, j).$$

The direct product of two diagonal matrices is also diagonal, and matrices $E_{p,q}$ and E_{p_1,q_1} consequently have the following characteristic roots for $b = c = 0$:

$$E_{p,q}\colon \ a^{p+l}d^{p-l}(\bar{a})^{q+m}(\bar{d})^{q-m} \quad \begin{pmatrix} l = -p, -p+1, \ldots, p-1, p \\ m = -q, -q+1, \ldots, q-1, q \end{pmatrix}$$

$$E_{p_1,q_1}\colon \ a^{p_1+l_1}d^{p_1-l_1}(\bar{a})^{q_1+m_1}(\bar{d})^{q_1-m_1}$$

$$\begin{pmatrix} l_1 = -p_1, -p_1+1, \ldots, p_1-1, p_1 \\ m_1 = -q_1, -q_1+1, \ldots, q_1-1, q_1 \end{pmatrix}$$

or, on observing that $ad = 1$:

$$E_{p,q}\colon \ a^{2l}(\bar{a})^{2m}; \ E_{p_1,p_1}\colon \ a^{2l_1}(\bar{a})^{2m_1}.$$

We can take any non-zero complex number for a, and it can clearly be chosen so that the set of characteristic roots of the $E_{p,q}$ differs from the set of roots of the E_{p_1,q_1}, which proves the non-equivalence of representations (187) for different choices of j and j'. We observe that, with $j' = 0$, representation (187) is the same as representation (185), whilst with $j=0$ it is the same as the representation obtained from (185) with $j = j'$ and a, b, c, d replaced by their conjugates. A singular feature of representations (187) may be noted. They are not equivalent to unitary representations. If they were, all the characteristic roots of any representation matrix would have to have unit moduli, whereas we saw above that the characteristic roots for representations $E_{p,q}$

with $b = c = 0$ are equal to $a^{2l}(\bar{a})^{2m}$ and can evidently have moduli differing from unity. The only exception is $E_{0,0}$, which is the trivial identity representation in which unity corresponds to each element of group (183).

We saw in [66] that, if a given representation, not necessarily equivalent to a unitary representation, is reducible, i.e. equivalent to a representation with quasi-diagonal matrices of the same structure, a matrix must exist which commutes with all the matrices of the representation and which is not a scalar matrix. Hence, to prove the irreducibility of any representation (187), we only need to show that any matrix commuting with all the matrices of (187) must be scalar. This can be done precisely as in [68]. Representations (187) are thus mutually non-equivalent, and each is irreducible. Use is often made of a definition of reducibility different from that of [58]: a representation is said to be reducible if all its linear transformations (say of order n) leave unchanged a subspace L_k, where $0 < k < n$.

We have seen [58] that if a representation reducible in this sense consists of unitary matrices, it is reducible in the sense of the definition of [65], i.e. it is equivalent to a quasi-diagonal representation. If a representation is not unitary, reducibility in the sense of the definition of [65] does not follow from the invariance of a certain subspace. It can be shown that every group representation (187) is not only irreducible in the sense that we have indicated, but it leaves no subspace unchanged. It can further be shown that every linear representation of group (183) is either equivalent to one of representations (187), or equivalent to a representation having a reduced formula and consisting of several of representations (187).

We saw in [73] that the composition of two linear group representations is equivalent to multiplying the objects of the representations. We can say in view of this that the objects of representations (187) are the expressions

$$\eta_{kk'} = \frac{x_1^{j+k} x_2^{j-k}}{\sqrt{(j+k)!\,(j-k)!}} \cdot \frac{y_1^{j'+k'} y_2^{j'-k'}}{\sqrt{(j'+k')!\,(j'-k')!}}$$
$$\begin{pmatrix} k = j,\ j-1,\ \ldots,\ -j+1,\ -j \\ k' = j',\ j'-1,\ \ldots,\ -j'+1,\ -j' \end{pmatrix},$$

where x_1 and x_2 undergo transformation (183) and y_1, y_2 undergo (184).

We have spoken so far of linear representations of the group consisting of positive Lorentz transformations [64]. The positive transformations form only a part of the Lorentz transformations with

unity determinant; in addition, there exist the Lorentz transformations with (-1) determinant. The study of the structure of these more general sets of transformations and the extension of linear representations of the positive Lorentz transformation group to the total Lorentz group presents certain special features by comparison with the group of orthogonal transformations in three-dimensional space. It must be noted that we can lay down the requirement, when defining the total Lorentz group, that the direction of measuring time remains unchanged. In this case, we must add reflection to the Lorentz group:

$$x_1' = -x_1; \; x_2' = -x_2; \; x_3' = -x_3; \; x_4' = x_4.$$

A discussion of all the points mentioned can be found, for instance, in Cartan's *Leçons sur la theorie des spineurs* (Hermann, Paris, 1938) and in Van der Waerden's *Die gruppentheoretische Methode in der Quantenmechanik* (Springer, Berlin, 1932).

81. Theorem on the simplicity of the Lorentz group. We now show, by using a method similar to that employed in [70], that the Lorentz group is simple. For this, we only need to prove that there are no normal subgroups for the group G of transformations (183) other than the subgroup consisting of matrices E and $(-E)$. Suppose there is such a subgroup H_1, containing the matrix

$$A = \begin{Vmatrix} a, & b \\ c, & d \end{Vmatrix} \qquad (ad - bc = 1),$$

differing from E and $(-E)$. We want to show that H_1 is the same as G. If H_1 contains a matrix B, it also contains all the matrices $U^{-1} BU$, where U is any matrix of G. On taking into account the basic result concerning the reduction of matrices to the canonical form, as also the fact that the determinant of the matrix U, reducing any given matrix to the canonical form, can always be taken equal to unity [27], we see that it is sufficient to show that H_1 contains in the first place matrices with any permissible different characteristic roots t and t^{-1}, where t is any complex number differing from zero and $(+1)$. We observe here that the product of the characteristic roots of a matrix of group G must be equal to unity. In the second place, H_1 must contain matrices E and $(-E)$, and furthermore, taking into account the case of equal characteristic roots and of a double elementary divisor, we have to show that H_1 also contains the matrices

$$\begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix} \text{ and } \begin{Vmatrix} -1, & 0 \\ 1, & -1 \end{Vmatrix}. \tag{188}$$

We take a variable matrix of group G:

$$X = \begin{Vmatrix} x, & y \\ z, & x \end{Vmatrix} \quad (x^2 - yz = 1)$$

and form the matrix

$$Y = A(XA^{-1}X^{-1})$$

which, as in [70], must appear in H_1. We obtain for the trace s of Y:

$$s = 2 + bz^2 + cy^2 - \left[(a-d)^2 + 2bc\right] yz.$$

Since A differs from E and $(-E)$, we cannot have simultaneously: $b = c = 0$ and $a = d$. Hence s is not constant, and on varying z and y, we can assign to s any complex values. The characteristic roots of Y are given by the quadratic equation

$$\lambda^2 - s\lambda + 1 = 0.$$

We can thus obtain arbitrary values t and t^{-1} for these roots, and consequently H_1 contains all the matrices with different characteristic roots and unity determinant. H_1 clearly also contains E, as also $(-E)$, which can be written as the product:

$$-E = [t, t^{-1}] = [-t^{-1}, -t],$$

where each factor belongs to H_1. Moreover, matrices (188) are readily written as the products of two matrices with unity determinant and with different characteristic roots, whence it follows that H_1 also contains these matrices. For

$$\begin{Vmatrix} 1, & 0 \\ 1, & 1 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{\beta}, & 0 \\ 0, & \beta \end{Vmatrix} \cdot \begin{Vmatrix} \beta, & 0 \\ \frac{1}{\beta}, & \frac{1}{\beta} \end{Vmatrix}$$

$$(\beta \neq 0 \text{ and } \pm 1)$$

$$\begin{Vmatrix} -1, & 0 \\ 1, & -1 \end{Vmatrix} = \begin{Vmatrix} \frac{1}{\beta}, & 0 \\ 0, & \beta \end{Vmatrix} \cdot \begin{Vmatrix} -\beta, & 0 \\ \frac{1}{\beta}, & -\frac{1}{\beta} \end{Vmatrix}.$$

We have now shown that H_1 must coincide with G, i.e. G has no normal subgroup except that consisting of E and $(-E)$, and the positive Lorentz transformation group is shown to be simple. Hence it follows as in [70] that the group cannot have homomorphic (not isomorphic) representations.

82. Continuous groups. Structural constants. The groups of rotations of three-dimensional space and of positive Lorentz transformations provide examples of infinite groups which depend on continuously varying parameters. The role of parameters can be played say by the Eulerian angles in the case of rotation groups. The groups consist in these cases of linear transformations, and the parametric dependence of the groups amounts to the parametric dependence of the elements of the matrices by which the linear transformations are defined. Groups of linear transformations are discussed below.

Let the matrix elements a_{ik} of the linear transformations forming a group G be functions of r real parameters $\alpha_1, \alpha_2, \ldots, \alpha_r$, and let certain conditions which we shall indicate next be fulfilled. Let the a_{ik} be single-valued functions of the α_s for all values of these parameters, sufficiently close to zero, and let the identity element of G, characterized by the conditions $a_{ik} = 0$ for $i \neq k$ and $a_{ii} = 1$, correspond to the zero values of the parameters $\alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$. Suppose further that definite values of the α_s, sufficiently close to zero, correspond to the elements of G neighbouring the identity element. The closeness of a group element to the identity element amounts to the fact that the elements a_{ik} of the corresponding matrices are near zero for $i \neq k$ and near unity for $i = k$. We have with these assumptions a one-to-one correspondence of elements of G in a certain neighbourhood of the identity element with points of a neighbourhood round the origin of the real r-dimensional space T_r. We shall consider later, not just this local one-to-one correspondence, but a one-to-one correspondence as a whole, in which to each element of G there corresponds a definite point belonging to a domain V of the space T_r containing the origin as an interior point, and conversely, to any point of V there corresponds a definite element of G. For the present, we only require the above local correspondence. Elements of G will be written G_α, G_β, G_γ etc., corresponding to parametric values α_s, β_s, γ_s $(s = 1, 2, \ldots, r)$. Taking the local view-point, the parameters must be fairly close to zero and the group elements fairly close to the identity element.

We consider a product of group elements:

$$G_\beta\, G_\alpha = G_\gamma\,.$$

The parameters γ_s characterizing the element G_γ obtained as a result of the above multiplication are single-valued functions of the α_s and β_s:

$$\gamma_s = \varphi_s(\beta_1, \beta_2, \ldots, \beta_r;\quad \alpha_1, \alpha_2, \ldots, \alpha_r)\,. \tag{189}$$

We take these functions to be continuous, with continuous derivatives up to the fourth order for all a_s and β_s fairly near zero.

The correspondence of the identity element to the zero values of the parameters gives us at once:

$$\begin{array}{l} \varphi_s(\beta_1, \beta_2, \ldots, \beta_r;\ 0, 0, \ldots, 0) = \beta_s; \\ \varphi_s(0, 0, \ldots, 0;\ a_1, a_2, \ldots, a_r) = a_s \end{array} \quad (s = 1, 2, \ldots, r), \qquad (190)$$

whence

$$\begin{array}{l} \dfrac{\partial \varphi_i}{\partial \beta_k} = \delta_{ik} \text{ for } a_s = 0; \\ \dfrac{\partial \varphi_i}{\partial \beta_k} = \delta_{ik} \text{ for } \beta_s = 0. \end{array} \quad (s = 1, 2, \ldots, r) \qquad (191)$$

The parameters a_s corresponding to the inverse element G_a^{-1} are evidently given by

$$\varphi_s(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_r;\ a_1, a_2, \ldots, a_r) = 0 \qquad (s = 1, 2, \ldots, r), \quad (192)$$

these equations being valid if all the a_s and $\tilde{a}_s$ are set equal to zero. The functional determinant of the left-hand sides of equations (192) in $\tilde{a}_s$ is, by (191), equal to unity for a_s and $\tilde{a}_s$ equal to zero. Hence, by the implicit function theorem, equations (192) define the $\tilde{a}_s$ as continuous functions for all a_s fairly near zero, the $\tilde{a}_s$ being zero for $a_s = 0$. We expand functions (189) in powers of a_s and β_s, using Maclaurin's formula, the expansion being carried out as far as the third order terms. We obtain, on taking into account (190) and (191):

$$\gamma_s = a_s + \beta_s + \sum_{i,k} a_{i,k}^{(s)} a_i \beta_k + \sum_{i,k,l} a_{i,k,l}^{(s)} a_i a_k \beta_l + \\ + \sum_{i,k,l} b_{i,k,l}^{(s)} a_i \beta_k \beta_l + \varepsilon^{(s)}, \qquad (193)$$

where the $a_{i,k}^{(s)}$, $a_{i,k,l}^{(s)}$ and $b_{i,k,l}^{(s)}$ are numerical coefficients, $\varepsilon^{(s)}$ is of not less than the fourth order of smallness with respect to a_s and β_s, and the summation is over i, k and l from 1 to r. The numbers

$$C_{ik}^{(s)} = a_{ik}^{(s)} - a_{ki}^{(s)} \qquad (s, i, k = 1, 2, \ldots, r) \qquad (194)$$

are known as the *structural constants* of group G for the parameters a_s chosen.

If we bring in new parameters a_s' instead of a_s:

$$a_s = \omega_s(a_1', a_2', \ldots, a_r') \qquad (s = 1, 2, \ldots, r),$$

so that (1) $\omega_s(0, 0, \ldots, 0) = 0$, (2) the equations written are uniquely soluble with respect to the a_s' for all a_s sufficiently near zero, and (3) the

functions ω_s have a sufficient number of derivatives, then the structural constants in the new parameters a_s' will be different.

It follows at once from definition (194) that

$$C_{ki}^{(s)} = -C_{ik}^{(s)}. \qquad (194_1)$$

The following further relationships between the structural constants can be proved by using (192) and the associative rule for multiplication of the group elements:

$$\sum_{s=1}^{r} (C_{is}^{(t)} C_{jk}^{(s)} + C_{js}^{(t)} C_{ki}^{(s)} + C_{ks}^{(t)} C_{ij}^{(s)}) = 0 \quad (i, j, k, t = 1, 2, \ldots, r). \quad (194_2)$$

These relationships will not be used and their proof is omitted.

We return to (193). With a_s and β_s sufficiently near zero, the γ_s will also be near zero. Taking (191) and the implicit function theorem into account, we can say that (193) are soluble with respect to the β_s in some neighbourhood of the origin of the space T_r:

$$\beta_s = \psi_s(\gamma_1, \gamma_2, \ldots, \gamma_r;\ a_1, a_2, \ldots, a_r) \quad (s = 1, 2, \ldots, r). \qquad (195)$$

We note here that the condition: $\beta_s = 0$ $(s = 1, 2, \ldots, r)$ is equivalent to the condition: $\gamma_s = a_s$ $(s = 1, 2, \ldots, r)$. We use (193) and (195) to form two square matrices $S(a_s)$ and $T(a_s)$ of order r with elements $S_{ik}(a_s)$ and $T_{ik}(a_s)$ depending on parameters a_s:

$$S_{ik}(a_s) = \left(\frac{\partial \gamma_i}{\partial \beta_k}\right)_{\beta_s = 0}; \quad T_{ik}(a_s) = \left(\frac{\partial \beta_i}{\partial \gamma_k}\right)_{\gamma_s = a_s}$$
$$(s, i, k = 1, 2, \ldots, r). \qquad (196)$$

On recalling the differentiation rule for functions of a function and calculating the derivative of γ_i with respect to γ_k or the derivative of β_i with respect to β_k, we get:

$$S(a_s)\, T(a_s) = E \quad \text{and} \quad T(a_s)\, S(a_s) = E, \qquad (197)$$

where E is the unit matrix of order r. It follows from (191) that $S(a_s)$ becomes the unit matrix for $a_s = 0$. In view of (197), it now follows that $T(a_s)$ has the same properties. The structural constants may readily be expressed in terms of the elements of these matrices, in fact:

$$C_{ik}^{(p)} = \left(\frac{\partial S_{pk}(a_s)}{\partial a_i} - \frac{\partial S_{pi}(a_s)}{\partial a_k}\right)_{a_s = 0} \qquad (198)$$

or

$$C_{ik}^{(p)} = \left(\frac{\partial T_{pi}(a_s)}{\partial a_k} - \frac{\partial T_{pk}(a_s)}{\partial a_i}\right)_{a_s = 0}. \qquad (199)$$

For we have from (193) and (196):

$$a_{ik}^{(p)} = \left(\frac{\partial \gamma_p}{\partial \alpha_i \partial \beta_k}\right)_{\alpha_s = \beta_s = 0} = \left(\frac{\partial S_{pk}(\alpha_s)}{\partial \alpha_i}\right)_{\alpha_s = 0}, \tag{200}$$

and we can write on interchanging subscripts i and k:

$$a_{ki}^{(p)} = \left(\frac{\partial S_{pi}(\alpha_s)}{\partial \alpha_k}\right)_{\alpha_s = 0}, \tag{201}$$

whence (198) follows immediately. We have further, on taking (197) into account:

$$\sum_{j=1}^{r} S_{pj}(\alpha_s)\, T_{jk}(\alpha_s) = \delta_{pk}.$$

We differentiate both sides with respect to α_i then set all the α_s equal to zero. Recalling that $S(\alpha_s)$ and $T(\alpha_s)$ become the unit matrix for $\alpha_s = 0$ $(s = 1, 2, \ldots, r)$, we get

$$\left(\frac{\partial S_{pk}(\alpha_s)}{\partial \alpha_i}\right)_{\alpha_s = 0} + \left(\frac{\partial T_{pk}(\alpha_s)}{\partial \alpha_i}\right)_{\alpha_s = 0} = 0,$$

i.e. by (200):

$$a_{ik}^{(p)} = -\left(\frac{\partial T_{pk}(\alpha_s)}{\partial \alpha_i}\right)_{\alpha_s = 0},$$

whence (199) follows as above. Expressions (193) define the basic group operation that gives the parameters γ_s corresponding to the product $G_\beta\, G_\alpha$ in terms of the parameters α_s and β_s of elements G_α and G_β. It is clear from (193) that, for α_s and β_s near zero, the group operation reduces as a first approximation to: $\gamma_s = \alpha_s + \beta_s$, so that the group is Abelian to a first approximation. If the group is strictly Abelian, we have

$$\varphi_s(\beta_1 \beta_2, \ldots, \beta_r; \alpha_1 \alpha_2, \ldots, \alpha_r) = \varphi_s(\alpha_1 \alpha_2, \ldots, \alpha_r; \beta_1 \beta_2, \ldots, \beta_r)$$
$$(s = 1, 2, \ldots, r)$$

and $a_{ik}^{(s)} = a_{ki}^{(s)}$ in expansions (193), i.e. all the structural constants vanish for an Abelian group. For groups of a general type, the second order terms in (193) produce a trend away from commutativeness, the trend being indicated by the presence of non-zero structural constants. An expansion of the parameters $\tilde{\alpha}_s$ corresponding to the element G_α^{-1} may readily be obtained by using (193). We do this by putting $\gamma_s = 0$ in (193) and replacing β_s by $\tilde{\alpha}_s$. We get by the ordinary rule for differentiating implicit functions:

$$\tilde{\alpha}_s = -\alpha_s + \sum_{i,k} a_{ik}^{(s)} \alpha_i \alpha_k + \varepsilon_1^{(s)},$$

where $\varepsilon_1^{(s)}$ is of at least the third order of smallness with respect to $a_1, a_2, \ldots, a_r$.

83. Infinitesimal transformations. Suppose we have, as above, a continuous group G of linear transformations of order n defined by parameters a_s ($s = 1, 2, \ldots, r$). We shall write G_a as above for the matrix of the transformation corresponding to parameters a_s, so that a linear transformation has the form

$$\mathbf{x} = G_a \mathbf{u}, \tag{202}$$

where $\mathbf{u}$ is any vector of the complex n-dimensional space R_n and $\mathbf{x}$ is the transformed vector. We bring in the operation of matrix differentiation: if the elements of a matrix A are differentiable functions of a parameter t, the derivative of A with respect to t is defined as the matrix whose elements are the derivatives of the elements of A, i.e.

$$\left\{\frac{\mathrm{d}A}{\mathrm{d}t}\right\}_{ik} = \frac{\mathrm{d}\{A\}_{ik}}{\mathrm{d}t},$$

We obtain partial derivatives if the elements of A depend on several variables.

Similarly, if the components of a vector $\mathbf{z}(z_1, z_2, \ldots, z_n)$ of the space R_n are differentiable functions with respect to t, $\mathrm{d}\mathbf{z}/\mathrm{d}t$ is defined as the vector with components $\mathrm{d}z_i/\mathrm{d}t$, i.e. differentiation of a vector amounts to differentiation of its components [II, 107].

We now introduce the so-called *infinitesimal transformations of a group G:*

$$I_k = \left(\frac{\partial G_a}{\partial a_k}\right)_{a_s=0} \qquad (k = 1, 2, \ldots, r). \tag{203}$$

The symbol I_k clearly denotes a matrix of order n with numerical elements.

We now return to (202) and let $\mathbf{u}$ be a fixed vector, i.e. its components are independent of the a_s. Obviously, the transformed vector will in general depend on the parameters, and we next derive the fundamental differential equations of this vector. For this, we apply to both sides of (202) the linear operation defined by matrix G_β:

$$G_\beta \mathbf{x} = G_\gamma \mathbf{u},$$

where $G_\gamma = G_\beta G_a$, and parameters γ_s are given in terms of a_s and β_s in accordance with the basic group operation (193). We differentiate both sides of the last expression with respect to β_p then put $\beta_s = 0$,

i.e. $\gamma_s = \alpha_s$. We obtain on using definition (203):

$$I_p \mathbf{x} = \sum_{j=1}^{r} \left[\frac{\partial (G_\gamma \mathbf{u})}{\partial \gamma_j}\right]_{\gamma_s = \alpha_s} \left(\frac{\partial \gamma_j}{\partial \beta_p}\right)_{\beta_s = 0}.$$

The first factor under the summation sign is evidently equal to the derivative of the right-hand side of (202) with respect to α_j, and if we recall notation (196), we can re-write this last equation as

$$I_p \mathbf{x} = \sum_{j=1}^{r} S_{jp}(\alpha_s) \frac{\partial \mathbf{x}}{\partial \alpha_j} \qquad (p = 1, 2, \ldots, r).$$

If we introduce the vectors

$$X\left(\frac{\partial \mathbf{x}}{\partial \alpha_1}, \frac{\partial \mathbf{x}}{\partial \alpha_2}, \ldots, \frac{\partial \mathbf{x}}{\partial \alpha_r}\right) \quad \text{and} \quad Y(I_1 \mathbf{x}, I_2 \mathbf{x}, \ldots, I_r \mathbf{x}),$$

we can now write the above as a linear transformation:

$$Y = S^*(\alpha_s) X,$$

where $S^*(\alpha_s)$ denotes the transposed matrix as usual. We find, on multiplying on the left by $T^*(\alpha_s)$ and taking (197) into account:

$$X = T^*(\alpha_s) Y,$$

or in the expanded form:

$$\frac{\partial \mathbf{x}}{\partial \alpha_p} = \sum_{j=1}^{r} T_{jp}(\alpha_s) I_j \mathbf{x} \qquad (p = 1, 2, \ldots, r). \tag{204}$$

We have for the components x_k of the vector $\mathbf{x}$ defined by (202):

$$\frac{\partial x_k}{\partial \alpha_p} = \sum_{j=1}^{r} T_{jp}(\alpha_s) \sum_{t=1}^{n} \{I_j\}_{kt} x_t \qquad \begin{pmatrix} k = 1, 2, \ldots, n \\ p = 1, 2, \ldots, r \end{pmatrix}, \tag{205}$$

where the $\{I_j\}_{kt}$ are the components of matrix I_j. We must add to equations (204) for $\mathbf{x}$ the initial condition, following at once from (202):

$$\mathbf{x}|_{\alpha_s = 0} = \mathbf{u}, \tag{206}$$

where $\mathbf{u}$ is an arbitrarily assigned vector. We observe that the $T_{jp}(\alpha_s)$ appearing in the coefficients of (204) may be found directly in accordance with group operation (193). Equations (204) lead us to relationships between the I_j,; these may be derived simply by writing down the condition that the second derivative of $\mathbf{x}$ with respect to α_p and α_q is independent of the order of differentiation.

We have from (204):

$$\frac{\partial^2 \mathbf{x}}{\partial a_p\, \partial a_q} = \sum_{j=1}^{r} \left(\frac{\partial T_{jp}(a_s)}{\partial a_q} I_j \mathbf{x} + T_{jp}(a_s) I_j \frac{\partial \mathbf{x}}{\partial a_q} \right)$$

or, on replacing $\partial \mathbf{x}/\partial a_q$ by its expression from (204) with $p = q$:

$$\frac{\partial^2 \mathbf{x}}{\partial a_p\, \partial a_q} = \sum_{j=1}^{r} \frac{\partial T_{jp}(a_s)}{\partial a_q} I_j \mathbf{x} + \sum_{j=1}^{r} \sum_{k=1}^{r} T_{jp}(a_s)\, T_{kq}(a_s)\, I_j I_k \mathbf{x}.$$

On interchanging p and q in the above and equating right-hand sides, we get the following corollary of system (205):

$$\left[\sum_{j=1}^{r} \left(\frac{\partial T_{jp}(a_s)}{\partial a_q} - \frac{\partial T_{jq}(a_s)}{\partial a_p} \right) I_j + \right.$$
$$\left. + \sum_{j=1}^{r} \sum_{k=1}^{r} \left(T_{jp}(a_s)\, T_{kq}(a_s) - T_{jq}(a_s)\, T_{kp}(a_s) \right) I_j I_k \right] \mathbf{x} = 0. \qquad (207)$$

We put all the a_s equal to zero in this relationship. We now have, on taking into account (199) and the fact that $T(a_s) = E$, when all the a_s vanish:

$$\left[\sum_{j=1}^{r} C_{pq}^{(j)} I_j + (I_p I_q - I_q I_p) \right] \mathbf{u} = 0,$$

whence, in view of the arbitrariness of vector $\mathbf{u}$, we have the following relationships between the infinitesimal transformations:

$$I_q I_p - I_p I_q = \sum_{j=1}^{r} C_{pq}^{(j)} I_j \qquad (p, q = 1, 2, \ldots, r). \qquad (208)$$

We have found the I_j and proved relationships (208) by starting from a given continuous group G and using equations (204). We show that this system, or what comes to the same thing, system (205), has a unique solution for the given initial condition (206). Suppose there are two solutions. In view of the linearity of (204), their difference must also satisfy the system and must become the null vector with $a_s = 0$. We thus want to show that the solution $\mathbf{x}$ of (204) with the zero initial condition is identically zero. To simplify the writing we shall take $r = 3$. Let our solution be $\mathbf{x}(a_1, a_2, a_3)$. We write (204) for $p = 1$ and put $a_2 = a_3 = 0$ on the right-hand side. We get an ordinary differential equation with a_1 as independent variable and the zero initial condition. The solution is identically zero by the familiar uniqueness theorem [II, 50], i.e. $\mathbf{x}(a_1, 0, 0) \equiv 0$. We now write (204) with

$p = 2$ and set $a_3 = 0$ on the right-hand side. This ordinary differential equation with a_2 as independent variable has the zero initial condition, as we have just shown: $\mathbf{x}(a_1, a_2, 0) = 0$ for $a_2 = 0$, and consequently, by the uniqueness theorem, $\mathbf{x}(a_1, a_2, 0) \equiv 0$. We now write (204) for $p = 3$. This ordinary differential equation has the zero initial condition $\mathbf{x}(a_1, a_2, a_3) = 0$ for $a_3 = 0$, and consequently $\mathbf{x}(a_1, a_2, a_3) \equiv 0$, which is what we wanted to prove.

Hence (204) can only lead to a single finite transformation (202) for given infinitesimal transformations I_j and given $T_{jp}(a_s)$, which are defined by group operation (193). In other words, the infinitesimal transformations define a group. This is essential as regards what follows. The proof of the existence of a solution of (204) is based on a general theorem for partial differential equations, which, as far as (204) is concerned, may be stated as follows: the necessary and sufficient condition for (204) to have a solution for any given initial condition (206) is that the square bracket in (207) vanishes identically with respect to a_s for any choice of p and q. We shall make no further use of this existence theorem.

84. Rotation groups. We take as an example the group of rotations of space about the origin. The corresponding third order matrices depend on three parameters. The role of parameters can be played say by the Eulerian angles. We shall now introduce different parameters a_1, a_2, a_3, in which all our future working will be performed. Any rotation may be considered as taking place about some axis l, passing through the origin, in a counter-clockwise direction and by an angle not exceeding π. Two rotations by an angle π about axes in opposite directions now lead to the same final position. We can thus look on any rotation as a vector from the origin in the direction of the axis of rotation and with a length equal to the angle of rotation. The projections (a_1, a_2, a_3) of this vector on the coordinate axes in fact serve as our parameters.

If we take a sphere V with centre at the origin and radius π and look on the ends of any diameter as identical, a one-to-one correspondence may be established between the points (a_1, a_2, a_3) of the sphere V and the elements of the rotation group. This applies not only in the neighbourhood of the origin and the identity element, but for the group as a whole, if we take the whole of the sphere V. All the matrices of the rotation group can be expressed in terms of parameters a_1, a_2, a_3 and the continuity and existence of derivatives mentioned above are satisfied.

We shall not deduce (193) for the basic group operation in the present case; instead, we determine the structural constants by evaluating directly the matrices of the infinitesimal transformations.

To evaluate I_1, we can take $a_2 = a_3 = 0$, differentiate the transformation matrix with respect to a_1, then set $a_1 = 0$. But for $a_2 = a_3 = 0$, we have a rotation about the x axis by the angle a_1, which leads us to the formulae:

$$x_1' = x_1$$
$$x_2' = x_2 \cos a_1 - x_3 \sin a_1,$$
$$x_3' = x_2 \sin a_1 + x_3 \cos a_1.$$

We obtain on differentiating the matrix of this transformation with respect to a_1 then setting $a_1 = 0$:

$$I_1 = \begin{Vmatrix} 0, & 0, & 0 \\ 0, & 0, & -1 \\ 0, & 1, & 0 \end{Vmatrix}. \tag{209}$$

Similarly,

$$I_2 = \begin{Vmatrix} 0, & 0, & 1 \\ 0, & 0, & 0 \\ -1, & 0, & 0 \end{Vmatrix}; \qquad I_3 = \begin{Vmatrix} 0, & -1, & 0 \\ 1, & 0, & 0 \\ 0, & 0, & 0 \end{Vmatrix}. \tag{210}$$

We can now evaluate the left-hand side of (208) and thus find the structural constants. Elementary working leads us to the following three relationships:

$$I_1 I_2 - I_2 I_1 = I_3; \quad I_2 I_3 - I_3 I_2 = I_1; \quad I_3 I_1 - I_1 J_3 = I_2. \tag{211}$$

If we expand the right-hand side of (202) in powers of a_s and confine ourselves to first order terms, we get

$$\mathbf{x} \doteq \mathbf{u} + (a_1 I_1 + a_2 I_2 + a_3 I_3)\,\mathbf{u}.$$

Hence $\mathbf{u}$ undergoes the following change as a result of the transformation:

$$\delta\mathbf{u} \doteq a_1 I_1 \mathbf{u} + a_2 I_2 \mathbf{u} + a_3 I_3 \mathbf{u}.$$

Each term on the right gives the change of $\mathbf{u}$ for a small rotation about one of the coordinate axes. For instance, we get the following change in the components (u_1, u_2, u_3) of $\mathbf{u}$ for a rotation by the small angle a_1 about the x axis:

$$\delta u_1 \doteq 0; \quad \delta u_2 \doteq -u_3 a_1; \quad \delta u_3 \doteq u_2 a_1.$$

Here, as above, we have confined ourselves to first order terms in a_1.

85. Infinitesimal transformations and representations of the rotation group. We next show the connection between the above discussion of infinitesimal transformations and representations of the rotation group. We shall assume a one-to-one representation in the neighbourhood of the identity transformation by matrices $F(a_1, a_2, a_3)$ of order n, the matrix elements being assumed continuous and differentiable functions of the parameters a_1, a_2, a_3. Any rotation D can be obtained as the product of a finite number of rotations of the above neighbourhood, and the product of the corresponding representation matrices gives the representation for D. The representation as a whole may be many-valued, however, since we can return to the initial rotation by continuous variation of the parameters and hence obtain a new representation for this rotation. We had this situation previously in the case of two-valued representations of the total rotation group [69].

We have the same group operation for matrices $F(a_1, a_2, a_3)$ as for the rotations themselves, and consequently the same structural constants. We can form infinitesimal transformations I_k for the group G' of matrices $F(a_1, a_2, a_3)$. The I_k will be nth order matrices connected by expressions (211). If the I_k can be found, we can write down differential equations (204) for a vector $\mathbf{x}$ of R_n, since the $T_{jp}(a_s)$ are defined solely by the group operation. These equations have a unique solution for a given initial condition (206), and this solution can clearly only be the transformation

$$\mathbf{x} = F\,(a_1, a_2, a_2)\,\mathbf{u}$$

which gives the representation of the rotation group in the neighbourhood of the identity element (transformation).

In the present case, $r = 3$, and on passing to the components of $\mathbf{x}$ in (204), we get $3n$ equations for the n components of

$$\mathbf{x}\,(x_1, x_2, \ldots, x_n)\,.$$

The only point of importance to us below is that (204) cannot have more than one solution for a given initial condition (206). As already mentioned, this can be stated as follows: *a representation of the rotation group is fully defined by its infinitesimal transformations* I_1, I_2, I_3.

It is thus entirely a question of determining the infinitesimal transformations of the representation, and we shall now consider this. We introduce new required matrices instead of I_1, I_2, I_3:

$$A_1 = -I_2 + iI_1; \quad A_2 = I_2 + iI_1; \quad A_3 = iI_3\,. \tag{212}$$

The following relationships are easily seen to hold between these, instead of (211):

$$\left.\begin{aligned} A_3 A_1 - A_1 A_3 &= A_1 \\ A_3 A_2 - A_2 A_3 &= -A_2 \\ A_1 A_2 - A_2 A_1 &= 2A_3. \end{aligned}\right\} \tag{213}$$

The representations by matrices $F(\alpha_1, \alpha_2, \alpha_3)$ must include, in particular, the representation of the Abelian group of rotations about the z axis, to the elements of which the matrices $F(0, 0, \alpha_3)$ correspond. All these matrices simultaneously take the diagonal form by a suitable choice of fundamental vectors, since irreducible representations of an Abelian group are of the first order. Transformation $F(0, 0, \alpha_3)$ becomes, for these fundamental vectors [69]:

$$F(0,\ 0,\ \alpha_3)\,\mathbf{v} = \mathrm{e}^{l\alpha_3}\,\mathbf{v}$$

or, on setting $l = -im$ and writing $\mathbf{v}_m$ for $\mathbf{v}$:

$$F(0,\ 0,\ \alpha_3)\,\mathbf{v}_m = \mathrm{e}^{-im\alpha_3}\,\mathbf{v}_m.$$

Since the condition that the representation be single-valued is laid down only in the neighbourhood of $\alpha_s = 0$ $(s = 1, 2, 3)$, we cannot assume that m is an integer. We obtain from the above, on the basis of the definition of I_3:

$$A_3\,\mathbf{v}_m = iI_3\,\mathbf{v}_m = i\left[\frac{\partial}{\partial\alpha_3} F(0,\ 0,\ \alpha_3)\ \mathbf{v}_m\right] = i\left(\frac{\partial}{\partial\alpha_3}\,\mathrm{e}^{-im\alpha_3}\,\mathbf{v}_m\right) = m\mathbf{v}_m\,.$$

Hence

$$A_3\,\mathbf{v}_m = m\mathbf{v}_m, \tag{214}$$

i.e. $\mathbf{v}_m$ is the eigenvector of the operator A_3, corresponding to the eigenvalue m. If there are several eigenvectors, $\mathbf{v}_m$ denotes one of them.

We now prove the following lemma:

LEMMA. *If $\mathbf{v}$ is the eigenvector of operator A_3, corresponding to the eigenvalue a, $A_1\mathbf{v}$, if it differs from zero, is also an eigenvector of A_3, corresponding to the eigenvalue $(a + 1)$, and similarly $A_2\mathbf{v}$ is an eigenvalue of A_3, corresponding to the eigenvalue $(a - 1)$.*

We have $A_3\mathbf{v} = a\mathbf{v}$ by hypothesis, whence by (213):

$$\begin{aligned} A_3(A_1\mathbf{v}) &= (A_1 A_3 + A_1)\,\mathbf{v} = A_1(A_3\mathbf{v}) + A_1\mathbf{v} = \\ &= A_1(a\mathbf{v}) + A_1\mathbf{v} = (a+1)\,A_1\mathbf{v} \end{aligned}$$

and similarly

$$A_3(A_2\mathbf{v}) = (a-1)\,A_2\mathbf{v}.$$

The number of different eigenvalues of A_3 is not greater than n. There will be one or several of these with a maximum real part; we call this value (or one of the values) j, and write $\mathbf{v}_j$ for the corresponding eigenvector. By our lemma, $A_1 \mathbf{v}_j$ must relate to the eigenvalue $(j+1)$, whereas, by definition of j, there is no such eigenvalue of A_3, so that we must have

$$A_1 \mathbf{v}_j = 0. \tag{215}$$

By the lemma, the vectors

$$\mathbf{v}_{j-1} = A_2 \mathbf{v}_j; \quad \mathbf{v}_{j-2} = A_2 \mathbf{v}_{j-1}; \ \ldots \tag{216}$$

relate, if not zero, to the eigenvalues $(j-1)$, $(j-2)$, ... of operator A_3. Vector sequence (216) must naturally lead to the null vector in the end, since the number of different eigenvalues for A_3 is not greater than n. We now prove the formula

$$A_1 \mathbf{v}_k = \varrho_k \mathbf{v}_{k+1} \qquad (k = j,\ j-1,\ j-2,\ \ldots), \tag{217}$$

where the ϱ_k are integers. This is true for $k = j$ by (215), in which case $\varrho_j = 0$, whilst the null vector can be taken, say, for $\mathbf{v}_{j+1}$. We now suppose that (217) is true for any of the k concerned, and show that it is true for $(k-1)$. We have by (213), (216) and (217):

$$\begin{aligned} A_1 \mathbf{v}_{k-1} &= A_1 (A_2 \mathbf{v}_k) = (A_2 A_1 + 2A_3) \mathbf{v}_k = \\ &= A_2 (A_1 \mathbf{v}_k) + 2A_3 \mathbf{v}_k = A_2 (\varrho_k \mathbf{v}_{k+1}) + 2k\mathbf{v}_k = (\varrho_k + 2k) \mathbf{v}_k . \end{aligned}$$

We remark that, with $k = j$, we make no use of the expression

$$A_2 \mathbf{v}_{k+1} = \mathbf{v}_k ,$$

because $\varrho_k = 0$ for $k = j$. Equation (217) is thus proved, and the numbers ϱ_k are defined by the relationships:

$$\varrho_{k-1} = \varrho_k + 2k; \quad \varrho_j = 0 \qquad (k = j,\ j-1,\ \ldots).$$

We obtain by calculating the values successively:

$$\varrho_k = j(j+1) - k(k+1),$$

i.e.

$$A_1 \mathbf{v}_k = [j(j+1) - k(k+1)] \mathbf{v}_{k+1} \qquad (k = j,\ j-1,\ \ldots). \tag{218}$$

We use these equations to find the subscript s of the first of vectors (216) that vanishes, i.e. $\mathbf{v}_s = 0$ and $\mathbf{v}_{s+1} \neq 0$. It follows from (217) that $\varrho_s = 0$, i.e.

$$j(j+1) - s(s+1) = 0 .$$

This is a quadratic equation in s with roots $s = j$ and $s = -(j+1)$. The value $s = j$ is unsuitable, since $\mathbf{v}_j$ is not zero and does not appear in sequence (216). Hence the vectors of (216):

$$\mathbf{v}_j, \mathbf{v}_{j-1}, \ldots, \mathbf{v}_{-j+1}, \mathbf{v}_{-j} \tag{219}$$

differ from zero, and $A_2 \mathbf{v}_{-j} = 0$. There are $(2j+1)$ of these vectors, whence it is clear that j is either a non-negative integer or half a positive odd integer. If $2j+1 = n$, we can take vectors (219) as the fundamental set in space R_n. On the other hand, if $2j+1 < n$, they form a subspace L_{2j+1} in R_n. Suppose that this latter case holds. Each $\mathbf{v}_k$ of sequence (219) satisfies the equation:

$$A_3 \mathbf{v}_k = k\mathbf{v}_k \quad (k = j, j-1, \ldots, -j+1, -j).$$

We have further, $A_2 \mathbf{v}_k = \mathbf{v}_{k-1}$, where $\mathbf{v}_{j-1} = 0$, together with (218). The operators A_1, A_2, A_3 displace subspace L_{2j+1} into itself, and the above formulae fully define the operators in the subspace. Moreover it follows at once from (216) and (218) that there is no subspace L_k inside L_{2j+1}, where $0 < k < 2j+1$, which remains unchanged on application of operators A_1, A_2, A_3. Having found the A_k, we can construct for the subspace L_{2j+1} the equations (204), which are necessarily satisfied by the vector

$$\mathbf{x} = F_j(a_1, a_2, a_3)\,\mathbf{u} \tag{220}$$

of the required representation in L_{2j+1}. This representation can leave no subspace L_k of L_{2j+1} invariant, i.e. it is irreducible in L_{2j+1}, since otherwise every A_s would have to leave L_k invariant, and this is untrue, as we have just seen. If $2j+1 = n$, the above discussion applies to R_n as a whole. With $2j+1 < n$, we have distinguished from the total representation in R_n a representation of order $(2j+1)$ which is irreducible in the sense indicated, i.e. it leaves invariant no subspace L_k for which $0 < k < 2j+1$. A direct consequence of our arguments is that there exists only one irreducible representation, leaving aside similar representations, of a given degree. Yet we have already formed irreducible unitary representations of any given degree in [69].

These in fact account for all the possible irreducible representations, and the representations based on operators A_s that we have constructed in L_{2j+1} must be similar to them.

Vectors (219) can be multiplied by an arbitrary non-zero numerical factor. In this case, numerical factors also make their appearance in (216) and (218). The factors can be chosen so that the following

relationships are finally obtained:

$$\left.\begin{aligned} A_1 \mathbf{v}_k &= \sqrt{j(j+1)-k(k+1)}\,\mathbf{v}_{k+1}, \\ A_2 \mathbf{v}_k &= \sqrt{j(j+1)-k(k-1)}\,\mathbf{v}_{k-1}, \\ A_3 \mathbf{v}_k &= k\mathbf{v}_k, \end{aligned}\right\} \tag{221}$$

where $\mathbf{v}_{j+1} = 0$ and $\mathbf{v}_{-j-1} = 0$.

We obtain with this choice of factors the same representations as were formed in [69] by starting from the quantities

$$\eta_l = \frac{x_1^{j+l} x_2^{j-l}}{\sqrt{(j+l)!\,(j-l)!}}\,. \tag{222}$$

The construction indicated above makes it possible to separate out the irreducible parts of any representation. All in all, it is a question of seeking the eigenvectors $\mathbf{v}_j$ of operator A_3 with the maximum eigenvalue and forming (216).

86. Representations of the Lorentz group. We take the group of linear transformations with unity determinant:

$$\begin{aligned} x_1' &= ax_1 + bx_2 \\ x_2' &= cx_1 + dx_2 \end{aligned} \quad (ad - bc = 1) \tag{223}$$

A transformation matrix contains four complex coefficients, with one relationship between them. Since three complex quantities remain arbitrary, we have six real parameters. We introduce these parameters by denoting a transformation matrix as:

$$A = \begin{Vmatrix} 1 + a_1 + ia_2, & a_3 + ia_4 \\ a_5 + ia_6, & d(a_s) \end{Vmatrix}, \tag{224}$$

where

$$d(a_s) = \frac{1 + (a_3 + ia_4)(a_5 + ia_6)}{1 + a_1 + ia_2}.$$

We obtain six infinitesimal transformations I_k, which are readily constructed. To find say I_1, we have to put all the a_s except a_1 equal to zero in A, differentiate A with respect to a_1, then set $a_1 = 0$. We find in this way that

$$I_1' = \begin{Vmatrix} 1, & 0 \\ 0, & -1 \end{Vmatrix}; \quad I_2' = \begin{Vmatrix} i, & 0 \\ 0, & -i \end{Vmatrix}; \quad I_3' = \begin{Vmatrix} 0, & 1 \\ 0, & 0 \end{Vmatrix};$$

$$I_4' = \begin{Vmatrix} 0, & i \\ 0, & 0 \end{Vmatrix}; \quad I_5' = \begin{Vmatrix} 0, & 0 \\ 1, & 0 \end{Vmatrix}; \quad I_6' = \begin{Vmatrix} 0, & 0 \\ i, & 0 \end{Vmatrix}.$$

The structural constants $C_{pq}^{(j)}$ appearing in (208) must be real by definition, and they can be determined from the relationships:

$$I'_p I'_q - I'_q I'_p = \sum_{j=1}^{6} C_{pq}^{(j)} I'_j \quad (p < q;\ p, q = 1, 2, \ldots, 6).$$

It must be observed here that no linear (non-trivial) relationship with real coefficients exists between the matrices I'_j, so that the following fifteen relationships can be obtained:

$$\begin{array}{lll}
I'_1 I'_3 - I'_3 I'_1 = 2I'_3, & I'_1 I'_4 - I'_4 I'_1 = 2I'_4, & I'_2 I'_3 - I'_3 I'_2 = 2I'_4, \\
I'_1 I'_5 - I'_5 I'_1 = -2I'_5, & I'_1 I'_6 - I'_6 I'_1 = -2I'_6, & I'_2 I'_5 - I'_5 I'_2 = -2I'_6, \\
I'_3 I'_5 - I'_5 I'_3 = I'_1, & I'_3 I'_6 - I'_6 I'_3 = I'_2, & I'_4 I'_5 - I'_5 I'_4 = I'_2 \\
I'_2 I'_4 = I'_4 I'_2 = -2I'_3, & I'_1 I'_2 - I'_2 I'_1 = 0, & \\
I'_2 I'_6 - I'_6 I'_2 = 2I'_5, & I'_3 I'_4 - I'_4 I'_3 = 0, & \\
I'_4 I'_6 - I'_6 I'_4 = -I'_1, & I'_5 I'_6 - I'_6 I'_5 = 0. &
\end{array}$$

If I_k $(k = 1, 2, \ldots, 6)$ are infinitesimal transformations for any representation of the group in question, they will also be connected by fifteen relationships

$$I_p I_q - I_q I_p = \sum_{j=1}^{6} C_{pq}^{(j)} I_j$$

with the same coefficients $C_{pq}^{(j)}$. If we introduce the notations:

$$\begin{array}{lll}
I_3 + iI_4 = 2A_1; & I_5 + iI_6 = 2A_2; & I_1 + iI_2 = 4A_3; \\
I_3 - iI_4 = 2B_1; & I_5 - iI_6 = 2B_2; & I_1 - iI_2 = 4B_3,
\end{array} \tag{225}$$

the fifteen relationships may be written as follows:

$$A_p B_q - B_q A_p = 0 \quad (p,\ q = 1,\ 2,\ 3) \tag{226}$$

together with the six relationships:

$$\begin{array}{ll}
A_3 A_1 - A_1 A_3 = A_1, & \\
A_3 A_2 = A_2 A_3 = -A_2, & (227) \\
A_1 A_2 = A_2 A_1 = 2A_3, &
\end{array}
\qquad
\begin{array}{ll}
B_3 B_1 - B_1 B_3 = B_1, & \\
B_3 B_2 - B_2 B_3 = -B_2, & (228) \\
B_1 B_2 - B_2 B_1 = 2B_3. &
\end{array}$$

Notice that relationships (226) and (227) are satisfied trivially if we take the matrices I'_k, since in this case $A'_k = 0$ $(k = 1, 2, 3)$. Relationships (227) are the same as (213) and the arguments of the previous

section remain in force. We apply the relationships to the infinitesimal transformations of any linear representation of group (223). If $\mathbf{v}_j$ is the eigenvector of operator A_3 related to the maximum eigenvalue, there are $(2j+1)$ eigenvectors $\mathbf{v}_k$ $(k=j,\ j-1, \ldots -j+1,\ -j)$ of operator A_3 which are transformed by operators A_1, A_2, A_3 in accordance with (221), where $\mathbf{v}_{j+1} = 0$ and $\mathbf{v}_{-j-1} = 0$. Let $L_{(j)}$ be the subspace formed by all the eigenvectors of A_3 related to the eigenvalue j. We show that if $\mathbf{v}$ belongs to $L^{(j)}$, the vectors $B_q \mathbf{v}$ $(q = 1, 2, 3)$ also belong to $L^{(j)}$. For, by (226):

$$A_3(B_q\mathbf{v}) = B_q(A_3\mathbf{v}) = B_q(j\mathbf{v}) = jB_q\mathbf{v},$$

whence it follows that $B_q \mathbf{v}$ is the eigenvector of A_3 corresponding to the eigenvalue j (or is the null vector), i.e. $B_q \mathbf{v}$ belongs to $L^{(j)}$. We can repeat our arguments of [85] for $L^{(j)}$, replacing operators A_k by B_k. The vector sequence $\mathbf{v}_{jk'}$ $(k' = j', j' - 1, \ldots, -j' + 1, -j')$ can thus be formed in $L^{(j)}$, the vectors being transformed in accordance with (221) with j replaced by j' and A_k by B_k. On repeated application of the operator A_2, each $\mathbf{v}_{jk'}$ yields $(2j + 1)$ vectors $\mathbf{v}_{kk'}$ $(k = j, j - 1, \ldots, -j + 1, -j)$. We thus finally obtain $(2j + 1)(2j' + 1)$ vectors $\mathbf{v}_{kk'}$, for which the following relationships hold:

$$\left.\begin{aligned}
A_1\mathbf{v}_{kk'} &= \sqrt{j(j+1) - k(k+1)}\,\mathbf{v}_{k+1,\,k'},\\
A_2\mathbf{v}_{kk'} &= \sqrt{j(j+1) - k(k-1)}\,\mathbf{v}_{k-1,\,k'},\\
A_3\mathbf{v}_{kk'} &= k\mathbf{v}_{kk'},\\
B_1\mathbf{v}_{kk'} &= \sqrt{j(j+1) - k'(k'+1)}\,\mathbf{v}_{k,\,k'+1},\\
B_2\mathbf{v}_{kk'} &= \sqrt{j(j+1) - k'(k'-1)}\,\mathbf{v}_{k,\,k'-1},\\
B_3\mathbf{v}_{kk'} &= k'\mathbf{v}_{kk'}.
\end{aligned}\right\} \tag{229}$$

These expressions define operators A_p and B_q in a $(2j + 1)(2j' + 1)$-dimensional space, and operators I_k are defined in accordance with (225), after which equations (204) can lead only to a single linear representation of the group. This is the representation that we formed in [80].

We have followed in these last sections the lines of the treatment in van der Waerden's *Die gruppentheoretische Methode in der Quantenmechanik* (Springer, Berlin, 1932).

87. Auxiliary formulae. We return to the formulae of [82]. We have

$$G_\beta G_\alpha = G_\gamma, \tag{230}$$

the γ_s being given in terms of α_s and β_s in accordance with (189) or (193), which

define the basic group operation. We form a matrix which we denote by $S(G_\beta, G_\alpha)$, depending on variables α_s and β_s, i.e. on the group elements G_α and G_β; the elements of the matrix are given by the following formulae:

$$S_{ik}(G_\beta, G_\alpha) = \frac{\partial \gamma_i}{\partial \beta_k} \qquad (i, k = 1, 2, \ldots r). \tag{231}$$

We have already considered this matrix in [82] for $\beta_s = 0$, i.e. with $G_\beta = E$, where E is the identity element of the group. Let us investigate the properties of the matrix. An immediate consequence of the definition is:

$$S(G_\beta, E) = I. \tag{232}$$

We show that also:

$$S(G_\beta, G_\alpha) : S(E, G_\beta) = S(E, G_\beta G_\alpha). \tag{233}$$

We put $G_\alpha = G_{\alpha''} G_{\alpha'}$, so that

$$G_\gamma = G_\beta G_\alpha = (G_\beta G_{\alpha''}) G_{\alpha'} = G_\delta G_{\alpha'} \qquad (G_\delta = G_\beta G_{\alpha''}).$$

We use the rule for differentiating functions of a function:

$$\frac{\partial \gamma_i}{\partial \beta_k} = \sum_{s=1}^{r} \frac{\partial \gamma_i}{\partial \delta_s} \cdot \frac{\partial \delta_s}{\partial \beta_k} = \sum_{s=1}^{r} S_{is}(G_\delta, G_{\alpha'}) S_{sk}(G_\beta, G_{\alpha''}),$$

whence

$$S(G_\beta, G_{\alpha''} G_{\alpha'}) = S(G_\delta, G_{\alpha'}) S(G_\beta, G_{\alpha''}).$$

If we put $G_\beta = E$, $G_{\alpha''} = G_\beta$ and $G_{\alpha'} = G_\alpha$ in this equation, we obtain (233). With $G_\alpha = G_\beta^{-1}$, we get an expression for the inverse matrix to $S(E, G_\beta)$:

$$S^{-1}(E, G_\beta) = S(G_\beta, G_\beta^{-1}). \tag{234}$$

Matrix $S(E, G_\beta)$ becomes $S(\beta_s)$ in the notation of [82], and the inverse matrix becomes $T(\beta_s)$. We shall write these at the moment as $S(G_\beta)$ and $T(G_\beta)$:

$$S(E, G_\beta) = S(G_\beta); \quad S^{-1}(E, G_\beta) = T(G_\beta). \tag{235}$$

We have

$$S(G_\beta) T(G_\beta) = T(G_\beta) S(G_\beta) = E. \tag{236}$$

Equation (233) gives

$$S(G_\beta, G_\alpha) = S(E, G_\gamma) S^{-1}(E, G_\beta) = S(G_\gamma) S^{-1}(G_\beta), \tag{237}$$

and (231) can be written in the form

$$\frac{\partial \gamma_i}{\partial \beta_k} = \sum_{s=1}^{r} S_{is}(G_\gamma) T_{sk}(G_\beta). \tag{238}$$

On multiplying both sides by $T_{mi}(G_\gamma)$ and summing over i, we get by (236):

$$\sum_{i=1}^{r} T_{mi}(G_\gamma) \frac{\partial \gamma_i}{\partial \beta_k} = T_{mk}(G_\beta). \tag{239}$$

We differentiate (238) with respect to β_l:

$$\frac{\partial^2\gamma_i}{\partial\beta_k\partial\beta_l} = \sum_{s,\,p=1}^{r} \frac{\partial S_{is}(G_\gamma)}{\partial\gamma_p}\frac{\partial\gamma_p}{\partial\beta_l} T_{sk}(G_\beta) + \sum_{s=1}^{r} S_{is}(G_\gamma)\frac{\partial T_{sk}(G_\beta)}{\partial\beta_l},$$

whence, using expression (238) for $\partial\gamma_p/\partial\beta_l$:

$$\frac{\partial^2\gamma_i}{\partial\beta_k\partial\beta_l} = \sum_{s,\,p,\,q=1}^{r} \frac{\partial S_{is}(G_\gamma)}{\partial\gamma_p} S_{pq}(G_\gamma)\, T_{ql}(G_\beta)\, T_{sk}(G_\beta) + \sum_{s=1}^{r} S_{is}(G_\gamma)\frac{\partial T_{sk}(G_\beta)}{\partial\beta_l}.$$

On interchanging k and l on the right-hand side, using the independence of the left-hand side on the order of differentiation and interchanging the variables of summation s and q, we get

$$\sum_{s,\,p,\,q=1}^{r}\left[\frac{\partial S_{is}(G_\gamma)}{\partial\gamma_p} S_{pq}(G_\gamma) - \frac{\partial S_{iq}(G_\gamma)}{\partial\gamma_p} S_{ps}(G\gamma)\right] T_{ql}(G_\beta)\, T_{sk}(G_\beta) =$$
$$= -\sum_{s=1}^{r} S_{is}(G_\gamma)\left[\frac{\partial T_{sk}(G_\beta)}{\partial\beta_l} - \frac{\partial T_{sl}(G_\beta)}{\partial\beta_k}\right].$$

We multiply both sides by the product $S_{if}(G_\beta)\, S_{kg}(G_\beta)\, T_{hi}(G_\gamma)$ and sum over i, k and l from 1 to r. Taking (236) into account, we get the equivalent system of equations:

$$\sum_{i,\,p=1}^{r}\left[\frac{\partial S_{ig}(G_\gamma)}{\partial\gamma_p} S_{pf}(G_\gamma) - \frac{\partial S_{if}(G_\gamma)}{\partial\gamma_p} S_{pg}(G_\gamma)\right] T_{hi}(G_\gamma) =$$
$$= -\sum_{k,\,l=1}^{r} S_{lf}(G_\beta)\, S_{kg}(G_\beta)\left[\frac{\partial T_{hk}(G_\beta)}{\partial\beta_l} - \frac{\partial T_{hl}(G_\beta)}{\partial\beta_k}\right].$$

We easily pass from these equations to the above, by multiplying both sides by the product $T_{fl_1}(G_\beta)\, T_{kg_1}(G_\beta)\, S_{i_1h}(G_\beta)$ and summing over f, g, and h. In the latter system, the left-hand side depends only on γ_s and the right-hand side only on β_s. Hence, in view of the arbitrariness of G_a in (230) and the independence of β_s and γ_s, both sides of the last equation must be equal to the same constant, and in particular:

$$\sum_{k,\,l=1}^{r} S_{lf}(G_\beta)\, S_{kg}(G_\beta)\left[\frac{\partial T_{hk}(G_\beta)}{\partial\beta_l} - \frac{\partial T_{hl}(G_\beta)}{\partial\beta_k}\right] = C_{fg}^{(h)}.$$

We can write, with a change of subscripts:

$$\sum_{s,\,t=1}^{r} S_{ti}(G_a)\, S_{sk}(G_a)\left[\frac{\partial T_{ps}(G_a)}{\partial a_t} - \frac{\partial T_{pt}(G_a)}{\partial a_s}\right] = -C_{ik}^{(p)}. \tag{240}$$

If we put $G_a = E$ in this identity, i.e. $a_s = 0$ $(s = 1, \ldots, r)$ and use the fact that $S(E) = E$, we get

$$-C_{ik}^{(p)} = \left[\frac{\partial T_{pk}(G_a)}{\partial a_i} - \frac{\partial T_{pi}(G_a)}{\partial a_k}\right].$$

On comparing with (199) of [82], we see that $C_{ik}^{(p)}$ are the structural constants that we defined above. On multiplying both sides of (240) by $T_{il}(G_a)\, T_{km}(G_a)$ and summing over i and k, we get by (236):

$$\frac{\partial T_{pm}(G_a)}{\partial a_l} - \frac{\partial T_{pl}(G_a)}{\partial a_m} = - \sum_{i,k=1}^{r} C_{ik}^{(p)} T_{il}(G_a)\, T_{km}(G_a). \tag{241}$$

We return to (207) and (208). As we saw, (208) is obtained by equating the square bracket in (207) to zero with $a_s = 0$ $(s = 1, \ldots, r)$. It is easily shown by using (241) that *it follows from* (208) *that the square bracket in* (207) *in fact vanishes for any* a_s.

We write the second term of this bracket as

$$\sum_{j,k=1}^{r} T_{jp}T_{kq}I_jI_k - \sum_{j,k=1}^{r} T_{jq}T_{kp}I_jI_k,$$

the argument G_a of T being omitted. If we interchange j and k in the second term here, we get

$$\sum_{j,k=1}^{r} T_{jp}T_{kq}(I_jI_k - I_kI_j) = \sum_{j,k\,s=1}^{r} T_{jp}T_{kq}C_{qp}^{(s)}I_s.$$

On transforming the first term

$$\sum_{j=1}^{r}\left(\frac{\partial T_{jp}}{\partial a_q} - \frac{\partial T_{jq}}{\partial a_q}\right) I_j$$

in the bracket of (207) in accordance with (241), we at once get the same result but with the reverse sign. We consider, along with $S(G_\beta, G_a)$, the matrix $S'(G_\beta, G_a)$, the elements of which are given by

$$\frac{\partial \gamma_i}{\partial a_k} = S'(G_\beta, G_a). \tag{242}$$

We can prove, precisely as above, the expressions:

$$\left.\begin{aligned} S'(E, G_a) &= I, \\ S'(G_\beta, G_a, E) &= S'(G_\beta, G_a)\, S'(G_a, E), \\ S'^{-1}(G_a, E) &= S'(G_a^{-1}, G_a), \end{aligned}\right\} \tag{243}$$

which we shall require later.

88. The formation of groups with given structural constants. The present section is concerned with the general outlines of the problem of constructing a group operation and a group of linear transformations with given structural constants satisfying (194_1) and (194_2). The construction is based on the theorem previously mentioned from the theory of partial differential equations, which we now proceed to formulate.

Suppose we have the following system of partial differential equations

$$\frac{\partial z_i}{\partial x_k} = X_{ik}(x_1, \ldots, x_n; z_1, \ldots, z_m) \tag{244}$$

$$(i = 1, 2, \ldots, m; \ k = 1, 2, \ldots, n).$$

We use the system to write down the condition that

$$\frac{\partial^2 z_i}{\partial x_k \partial x_l} = \frac{\partial^2 z_i}{\partial x_l \partial x_k}.$$

It evidently has the form

$$\frac{\partial X_{ik}}{\partial x_l} + \sum_{s=1}^{m} \frac{\partial X_{ik}}{\partial z_s} \cdot \frac{\partial z_s}{\partial x_l} = \frac{\partial X_{il}}{\partial x_k} + \sum_{s=1}^{m} \frac{\partial X_{il}}{\partial z_s} \cdot \frac{\partial z_s}{\partial x_k},$$

or, substituting for $\partial z_s/\partial x_l$ and $\partial z_s/\partial x_k$ from (244):

$$\frac{\partial X_{ik}}{\partial x_l} + \sum_{s=1}^{m} \frac{\partial X_{ik}}{\partial z_s} \cdot X_{sl} = \frac{\partial X_{il}}{\partial x_k} + \sum_{s=1}^{m} \frac{\partial X_{il}}{\partial z_s} X_{sk} \qquad (k \neq l). \tag{245}$$

This equation expresses the relationship between variables x_k, z_l.

THEOREM. *If the X_{ik} are continuous functions and their partial derivatives that appear in* (245) *are continuous at and in the neighbourhood of $x_k = x_k^{(0)}$, $z_i = z_i^{(0)}$, and if all relationships* (245) *are satisfied identically with respect to x_k, z_i, then system* (244) *is uniquely soluble for the initial conditions*

$$z_i \Big|_{x_k = x_k^{(0)}} = z_i^{(0)}.$$

The satisfaction of all equations (245) as an identity with the continuity conditions mentioned is generally known as the condition for complete integrability of system (244). We now sketch out the construction of a group operation and of a group of linear transformations with given structural constants.

Let the structural constants $C_{ik}^{(p)}$, where $i, k, p = 1, 2, \ldots$, be given, and let (194_1) and (194_2) be satisfied by the constants.

We can verify by solving system (241) with respect to the partial derivatives that the relationships mentioned represent the condition for complete integrability of the system. Thus there exists a unique matrix $T(G_a)$ with elements $T_{pq}(G_a)$ $(p, q = 1, 2, \ldots, r)$ which becomes a unit matrix with $G_a = E$, i.e. with $a_s = 0$ $(s = 1, 2, \ldots, r)$ and which satisfies system (241). Having found $T(G_a)$, we can form its inverse $S(G_a) = T^{-1}(G_a)$. To obtain the group operation, we return to (238). The right-hand sides of these equations are known functions of β_s and γ_s $(s = 1, 2, \ldots, r)$. It can be verified that system (241) expresses the condition for complete integrability of system (238). Hence there exists a unique solution of system (238) which satisfies the initial conditions

$$\gamma_i \Big|_{\beta_k = 0} = a_i.$$

This solution in fact gives the group operation. The initial conditions express the fact that element G_γ, defined by (230), becomes G_a for $\beta_s = 0$ $(s = 1, 2, \ldots, r)$. We now pass to the formation of the group of linear transformations, i.e.

of the group of matrices of a given order, for the given structural constants, the matrix $T(G_a)$ being already obtainable as shown above. As we pointed out in [83], the condition for complete integrability of system (204) or (205) amounts to the vanishing identically of the square bracket in (207) for any choice of subscripts, whilst this last condition is fulfilled, as we proved in [87], if matrices I_s satisfy relationships (208). The solution of the problem must therefore begin with the construction of matrices I_s of a given order satisfying (208). This is a difficult algebraic problem. Having found the I_s, we can then assert that system (205) has a unique solution satisfying initial conditions (206). This solution in fact gives the matrix group with given structural constants $C_{ik}^{(p)}$.

It can be shown that integration of system (241) with initial conditions $T(E) = I$ amounts to integration of a system of ordinary linear differential equations with constant coefficients. We state the result. We form the system of ordinary linear differential equations with constant coefficients:

$$\frac{\mathrm{d}w_{ik}(t)}{\mathrm{d}t} = \delta_{ik} + \sum_{p,\,q=1}^{r} C_{pq}^{(i)} a_p \, w_{qk}(t)\,,$$

where $\delta_{ik} = 0$ for $i \neq k$, $\delta_{ii} = 1$ and $a_1, a_2, \ldots, a_r$ are assumed to be given constants. The functions $T_{ik}(a_s) = w_{ik}(1)$ now satisfy system (241) and initial conditions $T(E) = I$. A detailed treatment of the problem of forming a continuous group, given the structural constants, together with various other problems of the theory of continuous groups, may be found in L. S. Pontryagin's *Nepreryvnye gruppy* ("Continuous groups").

89. Integration over groups. We proved in [76, 77] a number of relationships which contained the sums of various quantities depending on the elements of a group, the summation being extended over all the group elements. In the case of a continuous group, the summation is replaced by integration with respect to the parameters defining the group elements. Let G be a continuous group such that there corresponds to it for some choice of parameters a bounded closed domain V (the domain with its boundary) in real r-dimensional space T_r defined by parameters $a_1, \ldots, a_r$, with a definite point of V corresponding to every element of G and vice versa. The functions $\varphi_j(\beta_1, \ldots, \beta_r; a_1, \ldots, a_r)$ defining the group operation are assumed continuous and differentiable a sufficient number of times inside V. Furthermore, these functions and their derivatives are assumed continuous up to and including the boundary of V. The parameters $\tilde{a}_s$, corresponding to the element G_a^{-1}, are assumed to be continuous functions of the parameters a_s. A group with these properties is generally said to be *compact*. To define integration over the group, we consider the determinant of matrix $S'(G_\beta, G_a)$ [87] and introduce the following notation for this:

$$\Delta'(G_\beta, G_a) = \left| \frac{\partial \gamma_i}{\partial a_k} \right|_1^r. \tag{246}$$

We have directly from (243):

$$\Delta'(E, G_a) = 1, \tag{247$_1$}$$

$$\Delta'(G_\beta G_a, E) = \Delta'(G_\beta, G_a) \cdot \Delta'(G_a, E). \tag{247$_2$}$$

Using the notation $\delta'(G_\beta) = \Delta'(G_\beta, E)$, we can write

$$\Delta'(G_\beta, G_\alpha) = \frac{\delta'(G_\beta G_\alpha)}{\delta'(G_\alpha)}. \tag{248}$$

Hence, observing that $\delta'(E) = \Delta'(E, E) = 1$, we get:

$$\Delta'(G_\alpha^{-1}, G_\alpha) = \frac{1}{\delta'(G_\alpha)}. \tag{249}$$

We introduce the further notation:

$$u'(G_\alpha) = \Delta'(G_\alpha^{-1}, G_\alpha). \tag{250}$$

In view of our above assumptions, $u'(G_\alpha)$ is also a continuous function in the domain V. It does not vanish, since

$$\frac{1}{u'(G_\alpha)} = \delta'(G_\alpha) = \Delta'(G_\alpha, E)$$

is also a continuous function. Observing that $u'(E) = 1$, we can say that $u'(G_\alpha)$ and $\delta'(G_\alpha)$ are positive functions. By (248), the same can be said of $\Delta'(G_\beta, G_\alpha)$.

Let $f(G_\alpha) = f(a_1, \ldots, a_r)$ be any function continuous in the closed domain V.

We define the integral of this function over group G by the formula:

$$\int_G f(G_\alpha)\, dG_\alpha = \int_V f(a_1 \ldots, a_r)\, u'(G_\alpha)\, da_1 \ldots da_r, \tag{251}$$

where we have the usual integral over domain V on the right-hand side. We show that this integral has the following property of left-hand invariance:

$$\int_G f(G_\alpha)\, dG_\alpha = \int_G f(G_\beta G_\alpha)\, dG_\alpha, \tag{252}$$

or in coordinate form:

$$\int_V f(a_1, \ldots, a_r)\, u'(G_\alpha)\, da \ldots da_r = \int_V f(\gamma_1, \ldots, \gamma_r)\, u'(G_\alpha)\, da_1, \ldots da_r, \tag{253}$$

where G_β is any fixed element of G. We replace the variable element G_α in the left-hand integral by the variable element G_δ, setting $G_\alpha = G_\beta\, G_\delta$, the domain of variation of parameters $\delta_1, \ldots, \delta_r$ being V as before. The transformation determinant is

$$\left| \frac{\partial a_i}{\partial \delta_k} \right|_1^r = \Delta'(G_\beta, G_\delta) = \frac{\delta'(G_\beta G_\delta)}{\delta'(G_\delta)} = \frac{u'(G_\delta)}{u'(G_\beta G_\delta)} = \frac{u'(G_\delta)}{u'(G_\alpha)},$$

and we get

$$\int_V f(a_1, \ldots, a_r)\, u'(G_\alpha)\, da_1 \ldots da_r = \int_V f(a_1, \ldots, a_r)\, u'(G_\alpha) \frac{u'(G_\delta)}{u'(G_\alpha)}\, d\delta_1 \ldots d\delta_r =$$
$$= \int_G f(G_\beta G_\delta)\, dG_\delta.$$

This is equivalent to (252). The replacement of G_α by G_δ on the right-hand side is of no importance.

Integrals invariant on the right are similarly formed. We introduce the determinant of the matrix $S(G_\beta, G_\alpha)$:

$$\Delta (G_\beta, G_\alpha) = \left| \frac{\partial \gamma_i}{\partial \beta_k} \right|_1^r. \tag{254}$$

We have as above:

$$\left.\begin{aligned} &\Delta (G_\alpha, E) = 1 \\ &\Delta (E, G_\beta G_\alpha) = \Delta (G_\beta, G_\alpha)\, \Delta (E, G_\beta) \\ &\Delta (G_\beta, G_\alpha) = \frac{\delta (G_\beta, G_\alpha)}{\delta (G_\beta)}, \end{aligned}\right\} \tag{255}$$

We introduce the positive function:

$$u (G_\alpha) = \frac{1}{\delta (G_\alpha)} = \Delta (G_\alpha, G_\alpha^{-1}), \tag{256}$$

and the integral is defined by the formula:

$$\int\limits_V f (\alpha_1, \dots , \alpha_r)\, u (G_\alpha)\, d\alpha_1 \dots d\alpha_r = \int\limits_G (G_\alpha)\, d\widetilde{G}_\alpha. \tag{257}$$

The tilde over the differential distinguishes the integral from (251).

We now have the property of right-hand invariance:

$$\int\limits_G f (G_\alpha)\, d\widetilde{G}_\alpha = \int\limits_G f(G_\alpha G_\beta)\, d\widetilde{G}_\alpha. \tag{258}$$

We now show that replacing G_α by G_α^{-1} under the sign of the integrand produces a transformation from a left to right-hand invariant integral, and conversely. We differentiate the equation $G_\lambda = G_\alpha\, G_\beta$, written parametrically, with respect to α_s, it being assumed throughout below that $G_\beta = G_\alpha^{-1}$:

$$\frac{\partial \lambda_i}{\partial \alpha_k} + \sum_{s=1}^{r} \frac{\partial \lambda_i}{\partial \beta_s} \cdot \frac{\partial \beta_s}{\partial \alpha_k} = 0, \text{ whence } \left| \frac{\partial \lambda_i}{\partial \alpha_k} \right|_1^r = (-1)^r \left| \frac{\partial \lambda_i}{\partial \beta_k} \right|_1^r \left| \frac{\partial \beta_i}{\partial \alpha_k} \right|_1^r,$$

so that we have, on taking (246) and (254) into account:

$$\left| \frac{\partial \beta_i}{\partial \alpha_k} \right|_1^r = (-1)^r \frac{\Delta(G_\alpha, G_\alpha^{-1})}{\Delta'(G_\alpha, G_\alpha^{-1})} = (-1)^r \frac{u(G_\alpha)}{u'(G_\alpha^{-1})}. \tag{259}$$

This determinant can be written in another form. It follows from the equation

$$\left| \frac{\partial \beta_i}{\partial \alpha_k} \right|_1^r \left| \frac{\partial \alpha_i}{\partial \beta_k} \right|_1^r = 1$$

that

$$\left| \frac{\partial \alpha_i}{\partial \beta_k} \right|_1^r = (-1)^r \frac{u'(G_\alpha^{-1})}{u(G_\alpha)} \tag{260}$$

or, on changing the places of G_α and G_α^{-1}:

$$\left| \frac{\partial \beta_i}{\partial \alpha_k} \right| = (-1)^r \frac{u'(G_\alpha)}{u(G_\alpha^{-1})}. \tag{261}$$

We now return to the integral. We find, on changing the variable of integration

in the usual way and using (260):

$$\int\limits_V f(a_1 \ldots, a_r)\, u'(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r = \int\limits_V f(\widetilde{\beta}_1, \ldots, \widetilde{\beta}_r)\, u'(G_a) \left|\left| \frac{\partial a_i}{\partial \beta_k} \right|\right|_1^r \mathrm{d}\beta_1 \ldots \mathrm{d}\beta_r =$$

$$= \int\limits_V f(\widetilde{\beta}_1, \ldots, \widetilde{\beta}_r)\, u'(G_a) \frac{u(G_\beta)}{u'(G_a)} \mathrm{d}\beta_1 \ldots \mathrm{d}\beta_r .$$

On cancelling $u'(G_a)$ and replacing the variable element G_β by the variable G_a, we get

$$\int\limits_V f(\widetilde{a}_1, \ldots, \widetilde{a}_r)\, u(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r = \int\limits_V f(a_1, \ldots, a_r)\, u'(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r . \tag{262}$$

We have similarly, on taking (261) into account:

$$\int\limits_V f(\widetilde{a}_1, \ldots, \widetilde{a}_r)\, u'(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r = \int\limits_V f(a_1, \ldots, a_r)\, u(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r . \tag{263}$$

We have made no use so far of the compactness of the group. The domain V may even be infinite, though we now have to assume that function $f(a_1, \ldots, a_r)$ is such that all the integrals written have a meaning. We use the compactness to show that $u(G_a) = u'(G_a)$. We consider for this the determinant

$$D(G_\beta, G_a) = \left| \frac{\partial \mu_i}{\partial \beta_k} \right|_1^r, \tag{264}$$

where $G_\mu = G^{-1}_{\iota}\, G_\beta\, G_a$, and we prove that

$$D(G_\beta, G_{a''} G_{a'}) = D(G^{-1}_{a''}\, G_\beta\, G_{a''}, G_{a'})\, D(G_\beta, G_{a''}) . \tag{265}$$

We can write:

$$G_\mu = (G_{a''} G_{a'})^{-1}\, G_\beta (G_{a''} G_{a'}) = G^{-1}_{a'}\, G_\gamma\, G_{a'} ,$$

where $G_\mu = G^{-1}_{a''}\, G_\beta\, G_{a''}$, so that

$$\left| \frac{\partial \mu_i}{\partial \beta_k} \right|_1^r = \left| \frac{\partial \mu_i}{\partial \nu_k} \right|_1^r \cdot \left| \frac{\partial \nu_i}{\partial \beta_k} \right|_1^r = D(G_\nu, G_{a'})\, D(G_\beta, G_{a''}) ,$$

whence (265) follows. On setting $G_\beta = E$ in (265), we get

$$D(E, G_{a''} G_{a'}) = D(E, G_{a'})\, D(E, G_{a''}) . \tag{266}$$

If we introduce the numerical function of an element:

$$\eta(G_a) = D(E, G_\sigma) , \tag{267}$$

we can write, by (266):

$$\eta(G_{a''} G_{a'}) = \eta(G_{a''})\, \eta(G_{a'}) , \tag{268}$$

i.e. corresponding to multiplication of elements we have multiplication of the corresponding values of the function $\eta(G_a)$. We clearly have:

$$\eta(E) = 1 \text{ and } \eta(G_a)\, \eta(G_a^{-1}) = 1 , \tag{269}$$

and $\eta(G_a)$ is continuous and positive in the closed domain V. We now show, using the compactness of the group, that $\eta(G_a) = 1$ for any element G_a. Sup-

pose, for a given G_a, we have $\eta(G_a) \neq 1$. If say $\eta(G_a) < 1$, then by (269): $\eta(G_a^{-1}) > 1$, and we can always assume that $\eta(G_a) > 1$. With this:

$$\eta(G_a^n) = [\eta(G_a)]^n \to \infty \text{ for } n \to \infty .$$

This contradicts the fact that the function $\eta(G_a)$, continuous in the closed domain V, must be bounded. We next consider the relationship between $u(G_a)$ and $u'(G_a)$. Let

$$G_\gamma = G_\beta G_a = G_a^{-1}(G_a G_\beta) G_a = G_a^{-1} G_\varrho G_a \quad (G_\varrho = G_a G_\beta) .$$

We have:

$$\left| \frac{\partial \gamma_i}{\partial \beta_k} \right|_1^r = \Delta(G_\beta, G_a) .$$

But on the other hand:

$$\left| \frac{\partial \gamma_i}{\partial \beta_k} \right|_1^r = \left| \frac{\partial \gamma_i}{\partial \varrho_k} \right|_1^r \left| \frac{\partial \varrho_i}{\partial \beta_k} \right|_1^r = D(G_\varrho, G_a)\, \Delta'(G_a, G_\beta),$$

i.e.

$$\Delta(G_\beta, G_a) = D(G_a G_\beta, G_a) \cdot \Delta'(G_a, G_\beta) .$$

We obtain on setting $G_\beta = G_a^{-1}$:

$$\Delta(G_a^{-1}, G_a) = D(E, G_a) \cdot \Delta'(G_a, G_a^{-1}) ,$$

i.e.

$$u(G_a^{-1}) = \eta(G_a)\, u'(G_a^{-1}) \quad \text{or} \quad u(G_a^{-1}) = u'(G_a^{-1})$$

for any G_a, since $\eta(G_a) = 1$. Thus for compact groups, the left invariant integral (251) is the same as the right invariant integral (257). Moreover, it follows from (262) and (263) that this integral is the same as

$$\int_V f(\tilde{a}_1, \ldots, \tilde{a}_r)\, u'(G_a)\, \mathrm{d}a_1 \ldots \mathrm{d}a_r .$$

The left and right invariant integrals may be different for non-compact groups. We take as an example the group of linear transformations of the form

$$z' = \mathrm{e}^{a_1} z + a_2 ,$$

where a_1 and a_2 vary from $(-\infty)$ to $(+\infty)$. Here, $r = 2$ and V is the whole plane. The composition of two transformations gives

$$z' = \mathrm{e}^{a_1} z + a_2; \quad z'' = \mathrm{e}^{\beta_1} z' + \beta_2; \qquad \gamma_1 = \varphi_1(\beta_1, \beta_2;\ a_1, a_2) = \beta_1 + a_1;$$

i.e.

$$z'' = \mathrm{e}^{\beta_1 + a_1} z + (\mathrm{e}^{\beta_1} a_2 + \beta_2), \qquad \gamma_2 = \varphi_2(\beta_1, \beta_2;\ a_1, a_2) = \mathrm{e}^{a_1} \beta_2 + a_2 .$$

Parameters $a_1 = a_2 = 0$ correspond to the identity element. The element G_a^{-1} has parameters $\tilde{a}_1 = -\tilde{a}_1$, $\tilde{a}_2 = -a_2 \mathrm{e}^{-a_1}$. We evaluate the functional determinants:

$$\Delta'(G_\beta, G_a) = \begin{vmatrix} 1, & 0 \\ 0, & \mathrm{e}^{a_1} \end{vmatrix} = \mathrm{e}^{a_1}; \quad \delta'(G_a) = \mathrm{e}^{a_1}; \quad u'(G_a) = \mathrm{e}^{-a_1};$$

$$\Delta(G_\beta, G_a) = \begin{vmatrix} 1, & \mathrm{e}^{a_1} \beta_2 \\ 0, & 1 \end{vmatrix} = 1; \quad \delta(G_a) = u(G_a) = 1 .$$

The left invariant integral has the form

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(a_1, a_2)\, e^{-a_1}\, da_1\, da_2$$

and the right invariant integral:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(a_1, a_2)\, da_1\, da_2\,.$$

We observe that requirements other than compactness of the group are possible in the proof of the equality of the right and left invariant integrals, i.e. of the equation $u(G_a) = u'(G_a)$.

Let G' be the subgroup consisting of elements of G of the form

$$G_\alpha\, G_\beta\, G_\alpha^{-1}\, G_\beta^{-1} \tag{270}$$

or of products of these elements, G_a and G_β being any elements of G.

It may readily be seen that, if an element G_γ is included among elements (270), so also is G_γ^{-1}.

Similarly, $G_\delta^{-1}\, G_\gamma\, G_\delta$, for any choice of G_δ of G, is likewise included among elements (270). It follows from what has been said above that G', the subgroup generated by elements (270), is a normal subgroup of G. The subgroup G' reduces to the identity element when and only when all the elements (270) are identity elements, i.e. when and only when G is an Abelian group. Subgroup G' may possibly be the same as G. This is true, in particular, if G is a non-Abelian simple group. Subgroup G' is generally known as the *derived group* of G.

It follows from the definition of (268) and (269) that $\eta(G_a\, G_\beta\, G_a^{-1}\, G_\gamma^{-1})_\beta = 1$, that $\eta(G_\gamma) = 1$ for all G_γ of G', and that $\eta(G_a)$ has the same value for all elements belonging to the same set with respect to G', i.e. $\eta(G_a)$ has the same value for any element of the group complementary to G'. If G' is the same as G, $\eta(G_a) = 1$ for any G_a of G. The same is true if the complementary group just mentioned is compact. But since

$$\eta(G_a) = 1,\ \text{we have}\ u(G_a) = u'(G_a)\,.$$

90. Orthogonality. Examples. The property of left and right invariance of the integral is analogous to the property in the case of finite groups that the product $G_s\, G_t$ or $G_t\, G_s$ of a variable element G_s and a fixed element G_t varies over all the group elements. We used this property to show that every group representation is equivalent to a unitary representation, and in proving the properties of orthogonality. Similar propositions can be proved for compact groups by using the invariant integral. If $A(G_a)$ are unitary matrices yielding an irreducible linear representation of a compact group G, and $B(G_a)$ are unitary matrices yielding a non-equivalent irreducible representation, we have the following expression for the orthogonality of the non-equivalent irreducible unitary representations, the matrix elements being denoted by two subscripts as usual:

$$\int_V \{A(G_a)\}_{ij}\, \{B(G_a)\}_{kl}\, u(G_a)\, da_1 \ldots da_r = 0\,. \tag{271}$$

We obtain for a single irreducible representation:

$$\int_V \{A(G_a)\}_{ij} \overline{\{A(G_a)\}}_{kl}\, u(G_a)\, da_1 \cdot da_r = \frac{\delta_{ik}\,\delta_{jl}}{p} \int_V u(G_a)\, da_1 \cdot da_r\,, \tag{272}$$

where p is the order of the matrices. Similarly, we have for the characters:

$$X(G_a) = \sum_{i=1}^{p} \{A(G_a)\}_{ii};\quad X'(G_a) = \sum_{i=1}^{q} \{B(G_a)\}_{ii}\,,$$

where p and q are the orders of matrices $A(G_a)$ and $B(G_a)$, and these have the following properties:

$$\int_V X(G_a)\, \overline{X'(G_a)}\, u(G_a)\, da_1 \cdot da_r = 0\,, \tag{273}$$

$$\int_V X(G_a)\, \overline{X(G_a)}\, u(G_a)\, da_1 \cdot da_r = \int_V u(G_a)\, da_1 \cdot da_r\,. \tag{274}$$

1. We now consider some examples. Let G be the Abelian group of rotations of the plane about the origin. Here, $r = 1$, and the single parameter is the angle of rotation a. We take a as belonging to the interval $(0,2\pi)$, the ends of this interval being regarded as identical. Successive rotations by angles a and β amount to a rotation of $\beta + a$, where the sum must always be made to belong to $(0,2\pi)$ by subtracting 2π if necessary. The functional determinants $\Delta(G_\beta, G_a)$ and $\Delta'(G_\beta, G_a)$ here reduce to the derivatives of $\beta + a$ with respect to β or a, i.e. to unity, so that $u(G_a) = u'(G_a) = 1$. We know that G has irreducible unitary representations of the first order e^{ima} $(m = 0, \pm 1, \pm 2, \ldots)$, and (273) and (274) give the familiar expressions:

$$\int_0^{2\pi} e^{im_1 a}\, \overline{e^{im_2 a}}\, da = \int_0^{2\pi} e^{i(m_1 - m_2) a}\, da = \begin{cases} 0 & \text{for } m_1 \neq m_2 \\ 2\pi & \text{for } m_1 = m_2\,. \end{cases} \tag{275}$$

We remark that, because of the need for making the sum $\beta + a$ belong to $(0,2\pi)$, we may have a singularity as regards the continuity and definition of the derivatives of the sum; this occurs when the sum is equal to 2π for a and β lying inside the interval.

2. We consider the group of rotations of three-dimensional space, using parameters rather different from those mentioned in [84]. Let the space be rotated by an angle ω about an axis forming angles a, β, γ, with the coordinate axes.

We introduce the four parameters:

$$a_0 = \cos\frac{1}{2}\omega;\ a_1 = \cos a \sin\frac{1}{2}\omega;\ a_2 = \cos\beta \sin\frac{1}{2}\omega;\ a_3 = \cos\gamma \sin\frac{1}{2}\omega\,. \tag{276}$$

These are connected by the relationship

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1\,. \tag{277}$$

The values $a_0 = 1$, $a_1 = a_2 = a_3 = 0$ correspond to the identity transformation. We can take a_1, a_2, a_3 as the parameters, and a_0 as a function of them.

If the rotations defined by parameters (a_0, a_1, a_2, a_3) and (b_0, b_1, b_2, b_3) are carried out successively, the parameters (c_0, c_1, c_2, c_3) of the resultant rotation may be readily seen to be given by

$$\begin{aligned} c_0 &= a_0 b_0 - a_1 b_1 - a_2 b_2 - a_3 b_3, & c_2 &= a_0 b_2 - a_1 b_3 + a_2 b_0 + a_3 b_1, \\ c_1 &= a_0 b_1 + a_1 b_0 + a_2 b_3 - a_3 b_2, & c_3 &= a_0 b_3 + a_1 b_2 - a_2 b_1 + a_3 b_0. \end{aligned} \tag{278}$$

We find from (277), treating a_0 as a function of a_1, a_2, a_3:

$$a_0 \frac{\partial a_0}{\partial a_j} + a_j = 0, \qquad (j = 1, 2, 3),$$

whence $\partial a_0 / \partial a_j = 0$ for E. Using this, we can easily form the functional determinant for $b_0 = 1$, $b_1 = b_2 = b_3 = 0$:

$$\frac{D(c_1, c_2, c_3)}{D(b_1, b_2, b_3)} = \begin{Vmatrix} a_0, & -a_3, & a_2 \\ a_3, & a_0, & -a_1 \\ -a_2, & a_1, & a_0 \end{Vmatrix} = a_0 (a_0^2 + a_1^2 + a_2^2 + a_3^2) =$$
$$= a_0 = \sqrt{1 - a_1^2 - a_2^2 - a_3^2}.$$

The invariant integral becomes

$$\int_V f(a_1, a_2, a_3) \frac{1}{\sqrt{1 - a_1^2 - a_2^2 - a_3^2}} \, da_1 \, da_2 \, da_3. \tag{279}$$

The domain V is the sphere with centre at the origin and unit radius. We remark that expressions (278) are obtained directly from the rule for multiplying quaternions:

$$c_0 + c_1 i + c_2 j + c_3 k = (a_0 + a_1 i + a_2 j + a_3 k)(b_0 + b_1 i + b_2 j + b_3 k),$$

the unities i, j and k being subject to the following multiplication rules:

$$i^2 = j^2 = k^2 = -1; \quad ij = -ji = k; \quad jk = -kj = i; \quad ki = -ik = j.$$

It is easy to establish the connection between parameters (a_0, a_1, a_2, a_3) and the Eulerian angles α, β, γ. The expressions are

$$a_0 = \cos \frac{1}{2} \beta \cos \frac{1}{2} (\alpha + \gamma); \qquad a_2 = \sin \frac{1}{2} \beta \sin \frac{1}{2} (\gamma - \alpha);$$

$$a_1 = \sin \frac{1}{2} \beta \cos \frac{1}{2} (\gamma - \alpha); \qquad a_3 = \cos \frac{1}{2} \beta \sin \frac{1}{2} (\alpha + \gamma).$$

The invariant integral may now be written in parameters (α, β, γ) as

$$\int_V f(\alpha, \beta, \gamma) \sin \beta \sin^2 \frac{1}{2} (\alpha - \gamma) \, d\alpha \, d\beta \, d\gamma, \tag{280}$$

where $0 \leqslant \alpha < 2\pi$, $0 \leqslant \beta < \pi$, $0 \leqslant \gamma < 2\pi$. We note that in integral (279) the function

$$\frac{1}{a_0} = \frac{1}{\sqrt{1 - a_1^2 - a_2^2 - a_3^2}}$$

becomes infinite if $\omega = \pi$. This is related to the fact that we have $\sin \omega/2$

instead of ω in expressions (276) for a_1, a_2, a_3. It is worth mentioning here that the properties discussed in [89] in connection with the definition of compactness only need to be fulfilled for a certain choice of parameters. The properties may well be lost on changing the parameters. Furthermore, the singularity in the continuity and definition of the derivatives mentioned at the end of the first example in connection with the group of rotations of the plane about the origin will also hold for the group of rotations of three-dimensional space.

It may also be remarked that the equality of the invariant integrals for the spatial rotation group has an immediate connection with the fact that this is a simple non-Abelian group.

3. It may easily be verified by direct evaluation that the left and right invariant integrals are equal for the Lorentz group which, as we have seen, is homomorphic with the group of linear transformations with unity determinant:

$$\left.\begin{aligned} x_1' &= a_0 x_1 + a_1 x_2 \\ x_2' &= a_2 x_1 + a_3 x_2 \end{aligned}\right\} \qquad (a_0 a_3 - a_1 a_2 = 1). \tag{281}$$

The values $a_0 = a_3 = 1$, $a_1 = a_2 = 0$ correspond to the identity element. We can take a_0 as a function of a_1, a_2, a_3 and take as parameters the real and imaginary parts of a_1, a_2, and $a_3 - 1$. The group operation amounts to multiplication of second order matrices, and we have:

$$c_0 = b_0 a_0 + b_1 a_2, \quad c_1 = b_0 a_1 + b_1 a_3, \quad c_2 = b_2 a_0 + b_3 a_2, \quad c_3 = b_2 a_1 + b_3 a_3. \tag{282}$$

If we put $a_k = a_k' + i a_k''$ $(k = 0, 1, 2, 3)$, the group parameters are a_1', a_2', a_3', a_1'', a_2'', a_3''. On writing further $b_k = \beta_k' + i\beta_k''$ and $c_k = \gamma_k' + i\gamma_k''$, to find the invariant integral we must first evaluate the functional determinants:

$$\frac{D(\gamma_1', \gamma_2', \gamma_3', \gamma_1'', \gamma_2'', \gamma_3'')}{D(\beta_1', \beta_2', \beta_3', \beta_1'', \beta_2'', \beta_3'')} \quad \text{for} \quad \beta_1' = \beta_2' = \beta_1'' = \beta_2'' = \beta_3'' = 0;\ \beta_3' = 1$$

or

$$\frac{D(\gamma_1', \gamma_2', \gamma_3', \gamma_1'', \gamma_2'', \gamma_3'')}{D(a_1', a_2', a_3', a_1'', a_2'', a_3'')} \quad \text{for} \quad a_1' = a_2' = a_1'' = a_2'' = a_3'' = 0;\ a_3' = 1,$$

the fact that $a_3' = 1$ and not zero for the identity transformation of the group being of no consequence. We get in both cases the same invariant integral:

$$\int_V f(a_1', a_2', a_3', a_1'', {}_2'', {}_3'',) \frac{1}{a_3'^2 + a_3''^2}\, da_1'\, da_2'\, da_3',\, da_1''\, da_2''\, da_3''. \tag{283}$$

The domain V is the total six-dimentional space. The equality of the invariant integrals is related to the fact that, for group (281), the sub-group G' formed by the multiple elements $G_\alpha G_\beta G_\alpha^{-1} G_\beta^{-1}$, to which we referred in [89], is the same as the group itself. For it is easily shown that G' does not reduce to the identity transformation or to the normal subgroup formed by elements E and $(-E)$. The actual working to find invariant integral (283) becomes simple on the basis of a lemma in which use is made of analytic functions of several complex variables (see Chap. IV of the second part of this volume).

Lemma. *Let $w_s = u_s + iv_s$ $(s = 1, 2, \ldots, k)$ be analytic functions of complex variables $z_s = x_s + iy_s$ $(s = 1, 2, \ldots, k)$. The functional determinant (Jacobian)*

of functions $(u_1, v_1, \ldots, u_k, v_k)$ *with respect to variables* $(x_1, y_1, \ldots, x_k, y_k)$ *is now equal to the square of the modulus of the functional determinant of functions* $(w_1, \ldots, w_k)$ *with respect to variables* $(z_1, \ldots, z_k)$.

We have (see Chapters I and IV of the second part of this volume):

$$\frac{\partial u_i}{\partial x_k} = \frac{\partial v_i}{\partial y_k}; \quad \frac{\partial v_i}{\partial x_k} = -\frac{\partial v_i}{\partial y_k},$$

and we can write:

$$\frac{D(u_1, v_1, \ldots, u_k, v_k)}{D(x_1, y_1, \ldots, x_k, y_k)} = \begin{vmatrix} a_{11}, b_{11}, & a_{12}, b_{12}, \ldots, & a_{1k}, b_{1k} \\ -b_{11}, a_{11}, & -b_{12}, a_{12}, \ldots, & -b_{1k}, a_{1k} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ a_{k1}, b_{k1}, & a_{k2}, b_{k2}, \ldots, & a_{kk}, b_{kk} \\ -b_{k1}, a_{k1}, & -b_{k2}, b_{k2}, \ldots, & -b_{kk}, a_{kk} \end{vmatrix},$$

where

$$a_{ik} = \frac{\partial u_i}{\partial x_k}; \qquad b_{ik} = \frac{\partial v_i}{\partial x_k}.$$

On adding to each odd column the next even column multiplied by i, we get the determinant:

$$\begin{vmatrix} c_{11}, b_{11}, & c_{12}, b_{12}, \ldots, & c_{1k}, b_{1k} \\ ic_{11}, a_{11}, & ic_{12}, b_{12}, \ldots, & ic_{1k}, a_{1k} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ c_{k1}, b_{k1}, & c_{k2}, b_{k2}, \ldots, & c_{kk}, b_{kk} \\ ic_{k1}, a_{k1}, & ic_{k2}, a_{k2}, \ldots, & ic_{kk}, a_{kk} \end{vmatrix} \quad (c_{ik} = a_{ik} + ib_{ik}).$$

Further, on subtracting from each even row the previous odd row multiplied by i, we get

$$\begin{vmatrix} c_{11}, b_{11}, & c_{12}, b_{12}, \ldots, & c_{1k}, b_{1k} \\ 0, \bar{c}_{11}, & 0, \bar{c}_{12}, \ldots, & 0, \bar{c}_{1k} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ c_{k1}, b_{k1}, & c_{k2}, b_{k2}, \ldots, & c_{kk}, b_{kk} \\ 0, \bar{c}_{k1}, & 0, \bar{c}_{k2}, \ldots, & 0, \bar{c}_{kk} \end{vmatrix}.$$

Transferring the odd columns to the left and the even rows to the top, we get

$$\begin{vmatrix} c_{11}, c_{12}, \ldots, c_{1k}, & b_{11}, b_{12}, \ldots, b_{1k} \\ c_{21}, c_{22}, \ldots, c_{2k}, & b_{21}, b_{22}, \ldots, b_{2k} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ c_{k1}, c_{k2}, \ldots, c_{kk}, & b_{k1}, b_{k2}, \ldots, b_{kk} \\ 0, \quad 0, \ldots, \quad 0, & \bar{c}_{11}, \bar{c}_{12}, \ldots, \bar{c}_{1k} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot \\ 0, \quad 0, \ldots, \quad 0, & \bar{c}_{k1}, \bar{c}_{k2}, \ldots, \bar{c}_{kk} \end{vmatrix},$$

whence it follows that

$$\frac{D(u_1 v_1, \ldots, u_k v_k)}{D(x_1 y_1, \ldots, x_k, y_k)} = \begin{vmatrix} c_{11}, \ldots, c_{1k} \\ \cdot\ \cdot\ \cdot \\ c_{k1}, \ldots, c_{kk} \end{vmatrix}, \begin{vmatrix} \bar{c}_{11}, \ldots, \bar{c}_{1k} \\ \cdot\ \cdot\ \cdot \\ \bar{c}_{k1}, \ldots, \bar{c}_{kk} \end{vmatrix} = \left| \frac{D(w_1, \ldots, w_k)}{D(z_1, \ldots, z_k)} \right|^2.$$

We next evaluate the function $u(G_a)$ in the invariant integral. For this, we have to evaluate, in accordance with the lemma, the functional determinant

$$\frac{D(c_1, c_2, c_3)}{D(b_1, b_2, b_3)} \quad \text{for} \quad b_0 = b_3 = 1\,;\; b_1 = b_2 = 0 \tag{284}$$

or

$$\frac{D(c_1, c_2, c_3)}{D(a_1, a_2, a_3)} \quad \text{for} \quad a_0 = a_3 = 1;\; a_1 = a_2 = 0. \tag{285}$$

It follows from the relationship $a_0 a_3 - a_1 a_2 = 0$ that:

$$-a_2 + a_3 \frac{\partial a_0}{\partial a_1} = 0; \qquad -a_1 + a_3 \frac{\partial a_0}{\partial a_2} = 0; \qquad a_0 + a_3 \frac{\partial a_0}{\partial a_3} = 0.$$

We have further:

$$\frac{\partial c_1}{\partial a_1} = b_0; \qquad \frac{\partial c_1}{\partial a_2} = 0; \qquad \frac{\partial c_1}{\partial a_3} = b_1,$$

$$\frac{\partial c_2}{\partial a_1} = b_2 \frac{\partial a_0}{\partial a_1}; \qquad \frac{\partial c_2}{\partial a_2} = b_2 \frac{\partial a_0}{\partial a_2} + b_3; \qquad \frac{\partial c_2}{\partial a_3} = b_2 \frac{\partial a_0}{\partial a_3},$$

$$\frac{\partial c_3}{\partial a_1} = b_2; \qquad \frac{\partial c_3}{\partial a_2} = 0; \qquad \frac{\partial c_3}{\partial a_3} = b_3,$$

whence (283) follows. The same result is obtained on the basis of expression (284).

INDEX

VOLUMES PUBLISHED IN THIS SERIES

Vol. 1. WALLACE—*Introduction to Algebraic Topology*
Vol. 2. PEDOE—*Circles*
Vol. 3. SPAIN—*Analytical Conics*
Vol. 4. MIKHLIN—*Integral Equations*
Vol. 5. EGGLESTON—*Problems in Euclidean Space: Application of Convexity*
Vol. 6. WALLACE—*Homology Theory on Algebraic Varieties*
Vol. 7. NOBLE—*Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*
Vol. 8. MIKUSINSKI—*Operational Calculus*
Vol. 9. HEINE—*Group Theory in Quantum Mechanics*
Vol. 10. BLAND—*The Theory of Linear Viscoelasticity*
Vol. 11. KURTH—*Axiomatics of Classical Statistical Mechanics*
Vol. 12. FUCHS—*Abelian Groups*
Vol. 13. KURATOWSKI—*Introduction to Set Theory and Topology*
Vol. 14. SPAIN—*Analytical Quadrics*
Vol. 15. HARTMAN and MIKUSINSKI—*Theory of Lebesgue Measure and Integration*
Vol. 16. KULCZYCKI—*Non-Euclidean Geometry*
Vol. 17. KURATOWSKI—*Introduction to Calculus*
Vol. 18. GERONIMUS—*Polynomials Orthogonal on a Circle and Interval*
Vol. 19. ELSGOLC—*Calculus of Variations*
Vol. 20. ALEXITS—*Convergence Problems of Orthogonal Series*
Vol. 21. FUCHS and LEVIN—*Functions of a Complex Variable, Volume II*
Vol. 22. GOODSTEIN—*Fundamental Concepts of Mathematics*
Vol. 23. KEENE—*Abstract Sets and Finite Ordinals*
Vol. 24. DITKIN and PRUDNIKOV—*Operational Calculus in Two Variables and its Applications*
Vol. 25. VEKUA—*Generalized Analytic Functions*
Vol. 26. AMIR-MOÉZ and FASS—*Elements of Linear Spaces*
Vol. 27. GRADSHTEIN—*Direct and Converse Theorems*
Vol. 28. FUCHS—*Partially Ordered Algebraic Systems*
Vol. 29. POSTNIKOV—*Foundations of Galois Theory*
Vol. 30. BERMANT—*Course of Mathematical Analysis, Part II*
Vol. 31. LUKASIEWICZ—*Elements of Mathematical Logic*
Vol. 32. VULIKH—*Introduction to Functional Analysis for Scientists and Technologists*
Vol. 33. PEDOE—*Introduction to Projective Geometry*